Development of novel polymorphic nuclear and chloroplast microsatellite markers in coast redwood (Sequoia sempervirens)

Journal:	*Plant Genetic Resources*
Manuscript ID	Draft
Manuscript Type:	Short Communication
Date Submitted by the Author:	n/a
Complete List of Authors:	Breidenbach, Natalie ; Georg-August University of Göttingen, Department of Forest Genetics and Forest Tree Breeding Gailing, Oliver ; Georg-August University of Göttingen, Department of Forest Genetics and Forest Tree Breeding Krutovsky, Konstantin; Georg-August University of Göttingen, Department of Forest Genetics and Forest Tree Breeding; Vavilov Institute of General Genetics, Russian Academy of Sciences, Laboratory of Population Genetics
Keywords:	cpSSR, EST-SSR, hexaploid
Abstract:	The range-wide genetic structure of the highly productive and valuable timber species Sequoia sempervirens (D. Don) Endl. is still insufficiently studied, although published data based on different genetic markers (nuclear and chloroplast microsatellites, AFLP, RFLP and isozymes) demonstrated relatively low population structure. However, more genetic markers are needed to increase the efficiency of population genetic studies in coast redwood. Therefore, we developed seven nuclear and five chloroplast microsatellite or simple sequence repeat (SSR) markers based on expressed sequence tags (ESTs) and complete chloroplast genome sequence, respectively. All selected markers were tested in a range-wide sample representing trees from 16 locations. They are highly polymorphic microsatellite loci with number of alleles ranging from 3 to 17, and number of effective alleles from 1.1 to 2.48. Coast redwood is a hexaploid species, and its chloroplasts are paternally inherited. Therefore, the chloroplast SSR (cpSSR) markers are especially useful for this species, because their genotyping is not affected by nuclear genome ploidy. Moreover, they showed high gene diversity for each locus within and across all populations and can be used to study range-wide population genetic structure, pollen based gene flow and long distance gene transfer. Coast redwood can propagate clonally, and nuclear polymorphic EST-SSRs can be used for clonal identification. They are linked with expressed genes and their variation can reflect variation in genes under selection, including those that could be potentially important for local adaptation of coast redwood considering the threat of climate change.

SCHOLARONE™
Manuscripts

1 **Short Communication**

2

3 Development of novel polymorphic nuclear and chloroplast 4 microsatellite markers in coast redwood (*Sequoia sempervirens*)

5

6 Natalie Breidenbach[1], Oliver Gailing[1], Konstantin V. Krutovsky[1,2,3,4*]

7

8 [1]*Department of Forest Genetics and Forest Tree Breeding, Georg-August University of*
9 *Göttingen, Büsgenweg 2, Göttingen 37077, Germany,* [2]*Laboratory of Population Genetics,*
10 *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow 119991,*
11 *Gubkina Str. 3, Russian Federation,* [3]*Laboratory of Forest Genomics, Genome Research and*
12 *Education Center, Institute of Fundamental Biology and Biotechnology, Siberian Federal*
13 *University, Akademgorodok 50a/2, Krasnoyarsk 660036, Russian Federation and*
14 [4]*Department of Ecosystem Sciences and Management, Texas A&M University, College*
15 *Station, Texas 77843-2138, USA*
16

17 [*]Corresponding author. Email: konstantin.krutovsky@forst.uni-goettingen.de

18 ORCID: http://orcid.org/0000-0002-8819-7084

19

20 Abstract

21 The range-wide genetic structure of the highly productive and valuable timber species
22 *Sequoia sempervirens* (D. Don) Endl. is still insufficiently studied, although published data
23 based on different genetic markers (nuclear and chloroplast microsatellites, AFLP, RFLP and
24 isozymes) demonstrated relatively low population structure. However, more genetic
25 markers are needed to increase the efficiency of population genetic studies in coast
26 redwood. Therefore, we developed seven nuclear and five chloroplast microsatellite or
27 simple sequence repeat (SSR) markers based on expressed sequence tags (ESTs) and
28 complete chloroplast genome sequence, respectively. All selected markers were tested in a
29 range-wide sample representing trees from 16 locations. They are highly polymorphic
30 microsatellite loci with number of alleles ranging from 3 to 17, and number of effective
31 alleles from 1.1 to 2.48. Coast redwood is a hexaploid species, and its chloroplasts are
32 paternally inherited. Therefore, the chloroplast SSR (cpSSR) markers are especially useful for
33 this species, because their genotyping is not affected by nuclear genome ploidy. Moreover,

For Peer Review

34 they showed high gene diversity for each locus within and across all populations and can be
35 used to study range-wide population genetic structure, pollen based gene flow and long
36 distance gene transfer. Coast redwood can propagate clonally, and nuclear polymorphic EST-
37 SSRs can be used for clonal identification. They are linked with expressed genes and their
38 variation can reflect variation in genes under selection, including those that could be
39 potentially important for local adaptation of coast redwood considering the threat of climate
40 change.
41
42 **Keywords**: coast redwood, *Sequoia sempervirens*, cpSSR, EST-SSR, microsatellites
43
For Peer Review

44 Introduction

45 The natural distribution of coast redwood (*Sequoia sempervirens* (D. Don) Endl.) extends
46 along the pacific coast of northern California to southern Oregon (Roy, 1966). It is an
47 important timber species, but there are only a few studies concerning its range-wide genetic
48 variation and differentiation (Hall and Langenheim, 1987; Brinegar, 2011; Douhovnikoff and
49 Dodd, 2011). All these studies found relatively low genetic differentiation among analysed
50 populations. On the contrary, fine scale and individual clone differentiations were highly
51 significant (Rogers, 2000; Douhovnikoff *et al.,* 2004; Ibañez *et al.,* 2009; Narayan *et*
52 *al.*, 2015). It seems that more genetic markers are needed to increase the efficiency of
53 population genetic studies in coast redwood. We used publicly available expressed sequence
54 tag (EST) and complete chloroplast genome sequence data to develop new microsatellite or
55 simple sequence repeat (SSR) markers for coast redwood.
56

57 Experimental

58 Needles from 309 trees in 16 locations were collected within the natural distribution range
59 of coast redwood in central and northern California (Online Supplementary 1). The collected
60 needles were dried on silica gel. DNA was extracted from two needles per sample using the
61 DNeasy 96 Plant Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions.
62 The extracted DNA was diluted in ddH_2O 1:10 for PCR and stored at -20 °C.
63 Possible microsatellite motifs were found by search for dinucleotide and trinucleotide
64 SSR-motifs for EST-SSRs in the transcriptome data of Scott *et al.* (2016) and for all SSR-motifs

65 in the complete chloroplast genome sequence (NCBI GenBank accession number
66 NC_030372.1) using the SciRoko program (Kofler *et al.,* 2007). Primer3 0.4.0 (Untergasser *et*
67 *al.,* 2012) was used to design the PCR primer pairs and oligoCalc (Kibbe, 2007) to check for
68 possible dimers and hairpins. In total, 57 primer pairs for different SSR-motifs were tested.
69 The EST-SSR forward primers were labelled with either Hex or 6-FAM (**Error! Reference**
70 **source not found.**). Following Schuelke (2000), PCRs for cpSSRs were performed with 5'
71 tailed 6-FAM dye-labelled M13 (5'-CACGACGTTGTAAACGAC-3') forward and PIG-tailed (5'-
72 GTTTCTT-3') reverse primers (Kubisiak *et al.,* 2013).
73 For all 12 primer pairs the following touch-down PCR program was used. First
74 denaturation at 94 °C for 15 min, followed by 17 cycles with denaturation at 94 °C for 1 min,
75 annealing at 67 °C for 1 min and elongation at 72 °C for 1 min, after each cycle the annealing
76 temperature was decreased by 1 °C, followed by 18 cycles with the annealing temperature
77 at 50 °C for 1 min.
78 PCR products were separated using the ABI genetic analyser 3130xl with GENSCAN ROX
79 500 as the internal size standard. GeneMapper 4.1. (Applied Biosystems) was used for
80 visualization of the PCR products.
81 We finally selected seven EST-SSRs that were genotyped following the procedure for
82 selecting reliable and consistently reproducible alleles suggested in Pfeiffer *et al*. (2011). In
83 total, 270 and 297 samples were genotyped for the EST-SSR and the cpSSR markers,
84 respectively. For each locus number of alleles (N_a), number of effective alleles (N_e), Shannon
85 Index (I), Nei's total gene diversity (H_t) and Nei's within population gene diversity (H_s) was
86 calculated.
87 For the EST-SSR markers N_a and I were calculated using the R-package "polysat" for
88 polyploid species (Clark *et al.,* 2011), and the converted binary input file was used to
89 calculate H_t and H_s with the PopGene (Yeh *et al.,* 1997) and N_e with the GenAlEx (Peakall and
90 Smouse, 2006; Peakall and Smouse, 2012) programs. The cpSSRs were analysed using
91 GenAlEx (N_a, N_e and I) and the R-package "hierfstat" (H_t and H_s) (Goudet and Jombart, 2015).
92

93 ## Discussion

94 The program SciRoKo identified 76 microsatellites with dinucleotide motif, six with
95 trinucleotide motif in ESTs, and six microsatellites with a dinucleotide motif in the
96 chloroplast genome. From all EST-SSRs one with trinucleotide and six with dinucleotide

For Peer Review

97 motifs and among six cpSSRs five dinucleotide motifs were reliable and polymorphic, and
98 thus were selected for further analysis (**Error! Reference source not found.**). The number of
99 alleles ranged from 5 to 17 for EST-SSRs and from 3 to 10 for cpSSRs, number of effective
100 alleles from 1.02 to 1.25 for EST-SSRs) and from 1.14 to 2.48 for cpSSRs), Shannon index
101 from 0.21 to 2.19 for EST-SSR and from 0.19 to 0.97 for cpSSRs, gene diversity from 0.02 to
102 0.19 for EST-SSRs and from 0.10 to 0.65 for cpSSRs (**Error! Reference source not found.**).
103 Due to the use of binary data to compute H_t and H_s for EST-SSRs, these parameters of gene
104 diversity cannot be directly compared between EST-SSRs and cpSSRs (Nybom 2004).
105 As expected, EST-SSRs showed lower number of alleles than the published random
106 nuclear SSRs (nSSRs) (Douhovnikoff and Dodd, 2011; Narayan *et al.,* 2015), which was
107 observed also earlier in other plants including conifers (e.g., Euyal *et al.,* 2001; Rungis *et al*.,
108 2004), supposedly because EST-SSRs are linked with expressed genes and can be under
109 selection (Bouk and Vision, 2007).
110 The already published cpSSR *Seq21* marker (Brinegar, 2011) showed more alleles and a
111 higher Shannon Index than the new cpSSRs, except for ss60974. Chloroplasts in coast
112 redwood are inherited paternally (Neale *et al*., 1989), and their markers are excellent tools
113 to study pollen based gene flow and its contribution to population similarity. CpSSRs are
114 usually highly polymorphic in conifer species (Vendramin *et al.,* 1999; Viard, 2001; Bucci *et*
115 *al.,* 2007), and we found similar results in coast redwood.
116 Coast redwood is a hexaploid species (Stebbin, 1948), which complicates microsatellite
117 genotyping (Douhovnikoff and Dodd, 2011; Narayan et al., 2015). The used microsatellite
118 scoring routine based on the genotype verification using ramets of known clones (Pfeiffer *et*
119 *al.,* 2011), resulted in unambiguous and reliable multilocus genotypes. Due to the usually
120 less polymorphic nature of EST-SSRs and the haploid nature of cpSSRs, both marker types
121 might be less prone to genotyping errors than nSSRs, which frequently have many alleles of
122 similar length (Hoffmann and Amos, 2005). Therefore, additional new EST-SSR and cpSSR
123 might increase the resolution power of microsatellite markers to study population structure
124 and local adaptation in coast redwoods. The highly polymorphic cpSSRs can be especially
125 useful for genotyping of individuals and clone assignment based on the specific haplotypes.
126

127 ## Acknowledgments

For Peer Review

128 We thank Alexandra Dolynska, Melanie Eckholdt and Babalola Jumoke for support during lab
129 work. This project is funded by the „Fachagentur für nachhaltige Ressourcen (FNR) des
130 Bundesministeriums für Ernährung und Landwirtschaft (BMEL)“.
131

132 ## References

133 Bouk A and Vision T (2007) The molecular ecologist's guide to expressed sequence tags.
134 *Molecular Ecology* 16: 907-924.

135 Brinegar C (2011) Rangewide genetic variation in coast redwood populations at a chloroplast
136 microsatellite locus. In R. B. Standiford, T. J. Weller, D. D. Piierto, and J. D. Stuart [eds.],
137 Proceedings of the Coast Redwood Forests in a Changing California: A Symposium for
138 Scientists and Managers, 241-249. USDA Forest Service General Technical Report PSW-
139 GTR-238. USDA Forest Service, Pacific Southwest Research Station, Albany, California,
140 USA.

141 Bucci G, Gonzalez-Martinez SC, Le Provost G, Plomion C, Ribeiro MM, Sebastiani F, Alía R and
142 Vendramin GG (2007) Range-wide phylogeography and gene zones in *Pinus pinaster* Ait.
143 Revealed by chloroplast microsatellite markers. *Molecular Ecology* 16: 2137-2153.

144 Clark L and Jasieniuk M (2011) polysat: An R package for polyploid microsatellite analysis.
145 *Molecular Ecology Resources* 11: 562-566.

146 Douhovnikoff V, Cheng AM and Dodd RS (2004) Incidence, size and spatial structure of
147 clones in second-growth stands of coast redwood, *Sequoia sempervirens* (Cupressaceae).
148 *American Journal of Botany* 91: 1140-1146.

149 Douhovnikoff V and Dodd RS (2011) Lineage divergence in coast redwood (*Sequoia*
150 *sempervirens*), detected by a new set of nuclear microsatellite loci. *The American Midland*
151 *Naturalist* 165: 22-37.

152 Euyal I, Sorrells M, Baum M, Wolters P and Powell W (2001) Assessment of genotypic
153 variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. *Euphytica*
154 119: 39-43.

155 Goudet J and Jombart T (2015) hierfstat: Estimation and Tests of Hierarchical F-Statistics. R-
156 package version 0.04-22. https://CRAN.R-project.org/package=hierfstat.

157 Hall GD and Langenheim JH (1987) Geographic variation in leaf monoterpenes of *Sequoia*
158 *sempervirens*. *Biochemical Systematics and Ecology* 15: 31-43.

159 Hoffmann JI and Amos W (2005) Microsatellite genotyping errors: detection approaches,
160 common sources and consequences for paternal exclusion. *Molecular Ecology* 14: 599-
161 612.

162 Ibañez MT, Caru M, Herrera MA, Gonzalez L, Martin LM, Miranda J and Navarro-Cerrillo RM
163 (2009) Clones identification of *Sequoia sempervirens* (D.Don) Endl. In Chile by using PCR-
164 RAPDs technique. *Journal of Zheijang University Science* B 10: 112-119.

165 Kibbe WA (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids*
166 *Research* 35: W43-W46.

167 Kofler R, Schlötterer C and Lelley T (2007) SciRoKo: a new tool for whole genome
168 microsatellite search and investigation. *Bioinformatics Application Notes* 23: 1683-1685.

169 Kubisiak TL, Nelson CD, Staton ME, Zhebentyayeva T, Smith C, Olukolu BA, Fang G-C, Hebard
170 FV, Anagnostakis S, Wheeler N, Sisco PH, Abbott AG and Sederoff RR (2013) A
171 transcriptome-based genetic map of Chinese chestnut (*Castanea mollissima*) and
172 identification of regions of segmental homology with peach (*Prunus persica*). *Tree*
173 *Genetics & Genomes* 9: 557-571.

174 Narayan L, Dodd RS and O'Hara KL (2015) A genotyping protocol for multiple tissue types
175 from the polyploidy tree species *Sequoia sempervirens* (Cupressaceae). *Applications in*
176 *Plant Science* 3: 1400110.

177 Neale D, Marshall K and Sederoff R (1989) Chloroplast and mitochondrial DNA are paternally
178 inherited in *Sequoia sempervirens* D.Don.Endl. *Proceedings of the National Academy of*
179 *Science* 86: 9347-9349.

180 Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific
181 genetic diversity in plants. *Molecular Ecology* 13: 1143-1155.

182 Peakall, R. and Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic
183 software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

184 Peakall, R. and Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic
185 software for teaching and research-an update. *Bioinformatics* 28: 2537-2539.

For Peer Review

186 Pfeiffer T, Roschanski AM, Pannell JR, Korbecka G and Schnittler M (2011) Characterization
187 of microsatellite loci and reliable genotyping in a polyploidy plant *Mercurialis perennis*
188 (Euphorbiaceae). *Journal of Heredity* 102: 479-488.

189 Rogers DL (2000) Genotypic diversity and clone size in old-growth populations of coast
190 redwood (*Sequoia sempervirens*). *Canadian Journal of Botany 78*: 1408-1419.

191 Rungis D, Bérubé Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J and Ritland K
192 (2004) Robust single sequence repeat markers for spruce (*Picea* spp.) from expressed
193 sequence tags. *Theoretical and Applied Genetics* 109: 1283-1294.

194 Schuelke M (2000) An economic method for the fluorescent labelling of PCR fragments.
195 *Nature Biotechnology* 18: 233-234.

196 Scott AD, Stenz NWM, Ingvarsson PK, Baum DA (2016) Whole genome duplication in coast
197 redwood (*Sequoia sempervirens*) and its implications for explaining the rarity of
198 polyploidy in conifers. *New Phytologist* 211: 186-193.

199 Stebbin GL Jr (1948) The chromosomes and relationships of *Metasequoia* and *Sequoia*.
200 *Science* 108: 95-98.

201 Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen SG (2012)
202 Primer3-new capabilities and interfaces. *Nucleic Acids Research* 40: e115.

203 Vendramin GG, Degen B, Petit RJ, Anzidei M Madaghiele A and Ziegenhagen B (1999) High
204 level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Molecular Ecology*
205 8: 1117-1126.

206 Viard F, El-Kassaby YA and Ritland K (2001) Diversity and genetic structure in populations of
207 *Pseudotsuga menziesii* (Pinaceae) at chloroplast microsatellite loci. *Genome* 44: 336-344.

208 Yeh FC, Yang R-C, Boyle T BJ (1997) POPGENE, the user-friendly shareware for population
209 genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta,
210 Canada.

For Peer Review

1 **Table 1**. Novel EST-SSR and cpSSR markers developed in coast redwood

SSR	Forward (F) and reverse (R) PCR primer 5'-3' sequence with fluorescent dye 6-FAM or HEX and M13 or PIG-tail adapters attached	Type	Repeat motif*	Allele size range, bp	N_a	N_e	I	H_t	H_s
ss36782	F: [FAM]-TCAGGGCAAAGCTAAAATCG R: CCAGGAAAGGAAAGGGAGAG	EST-SSR	$(GA)_{10}$	173-193	5	1.02	0.21	0.02	0.02
ss74800	F: [HEX]-GCATGACTCTGGTGGTGTTG R: GCAGCAGCCACTGTGAATAA	EST-SSR	$(TGG)_{8+1}$	205-238	12	1.13	1.48	0.09	0.08
ss91170	F: [FAM]-TCTGAAAAATGCCAAATCCA R: CGTGTGTCCTGTAAGTGCAAA	EST-SSR	$(CA)_{10+1}$	146-202	12	1.25	1.75	0.16	0.14
ss73361	F: [HEX]-AGGGTAGATGGGCGGTAGTT R: CGTCCGACAAGTTCAGTACG	EST-SSR	$(TC)_{9+1}$	190-216	14	1.24	2.19	0.16	0.14
ss73307	F: [HEX]-GAACTGTGAAAGCCCTTGGT R: GGGCGTGTTCTGTTTGAACT	EST-SSR	$(CA)_9$	208-235	17	1.10	1.59	0.07	0.06
ss73978	F: [HEX]-CCTGCAAACAATTCCAGCTT R: AGTGGGAATTATGGGGTTGG	EST-SSR	$(TC)_{10}$	214-220	5	1.11	0.66	0.09	0.08
ss114481	F: [FAM]-GGGTCAAGCGTGGTTATTGT R: TCTGGCATGATCCAAGTGTT	EST-SSR	$(TA)_9$	184-205	10	1.21	1.71	0.16	0.13
ss40585	F: [M13]-TCTTTTTCTTCAAGCACTTGTTTTT R: [PIG-TAIL]-TCAATCTACACGGGGATGTTT	cpSSR	$(AT)_8$	269-279	6	1.15	0.20	0.11	0.11
ss49836	F: [M13]- TGAAAGCTCTCGTGCGTATT R: [PIG-TAIL]-AGTTGAGTTCCCGGTTCTCC	cpSSR	$(AT)_{11}$	232-250	10	1.76	0.65	0.40	0.37
ss60974	F: [M13]- GCTCCGGCGTATAGAGAGG R: [PIG-TAIL]-GAGATTCCAATGGCTTTTGC	cpSSR	$(AT)_{12}$	221-237	9	2.48	0.97	0.65	0.53
ss85281	F: [M13]- TGACCATAGGTTCCTTCCTTTTT R: [PIG-TAIL]-TTCCGTTCCTTTCCATTTTG	cpSSR	$(AT)_8$	214-222	5	1.39	0.40	0.40	0.23
ss109990	F: [M13]-AAAAATCGACCGGATCACAA R: [PIG-TAIL]-TTCAAATAATAGAATGGAAAAACCAA	cpSSR	$(TA)_{11}$	224-232	3	1.14	0.19	0.10	0.09

2 *Number means number of repeat motifs in the original sequence; N_a: number of alleles; N_e: number of effective alleles; I: Shannon Index; H_t: total gene
3 diversity; H_S: gene diversity within population.

For Peer Review