GENOTYPIC DIVERSITY, PHYLOGEOGRAPHY, AND POPULATION STRUCTURE OF TWO LINEAGES OF *PHAEOCRYPTOPUS GAEUMANNII* IN THE PACIFIC NORTHWEST

***Patrick Bennett*[1] and *Jeffrey Stone*[1]**

INTRODUCTION

The ascomycete fungus *Phaeocryptopus gaeumannii* is the causal organism of Swiss needle cast (SNC), a foliage disease of Douglas-fir (*Pseudotsuga menziesii)*. The disease was first described in forest plantations in Europe in the 1920s (Boyce 1940). It has since been determined that *P. gaeumannii* is endemic to western North America and the native range of Douglas-fir in the northwestern United States. Initial assessments of genetic diversity and population structure performed using molecular markers known as single strand conformation polymorphisms (SSCPs) indicated the presence of two distinct sympatric populations (lineages) of the fungus in the Pacific Northwest (Winton *et al*. 2006). Subsequently a set of more variable molecular markers, microsatellites or short sequence repeats (SSRs), that permit more fine-scale analysis of *P. gaeumannii* populations were developed by Winton *et al*. (2007), but intensive sampling and analyses using these markers were not undertaken until now.

RESEARCH OBJECTIVES

The objectives of this study were to test whether *P. gaeumannii* comprises two distinct lineages in western Oregon and Washington, as inferred from SSCP data, and to compare the distributions of *P. gaeumannii* genotypes in the PNW with spatial variation in SNC severity as observed in ODF/USFS aerial surveys. The longer-term objectives of this research aim to assess *P. gaeumannii* population structure and diversity through the analysis of multi-locus microsatellite genotypes.

In: Ramsey, A. & P. Palacios (Comps). Proceedings of the 63rd Annual Western International Forest Disease Work Conference; 2015 Sept. 21-15; Newport, OR. [1]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon.

METHODS

Field Sampling

The 14 sites included in this analysis (Figure 1) were selected from a long-term monitoring plot network established by the Oregon State University Swiss Needle Cast Cooperative (SNCC) (Ritokova *et al*. 2013). Secondary branches were collected from the upper canopies of 5 randomly selected trees at each site. Additional foliage samples were collected from the lower and middle canopies of one of these trees at each site.

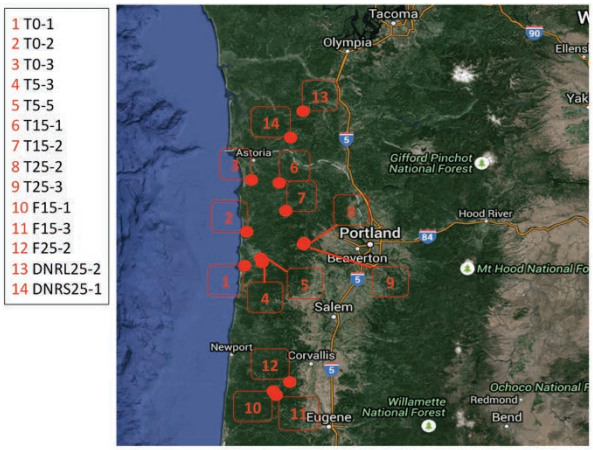


Figure 1. *Map showing the distribution of the 14 sites included in this analysis.*

Isolations and Culturing

Several needles bearing pseudothecia were attached to the covers of Petri dish with double-sided adhesive tape and suspended above the media to discharge ascospores onto the agar surface. After 24 hours the plates were examined under the stereomicroscope and single ascospores were isolated with the aid of flame-sterilized forceps onto 2% malt agar in 60 mm Petri dishes. These cultures were incubated at 20° C for 2–6 months to obtain adequate growth for DNA extraction.

Molecular Techniques

Genomic DNA was extracted by using the Qiagen DNeasy Plant kit following the manufacturer's protocol, with the addition of an initial maceration step. Ten microsatellite loci were amplified in 3 multiplexed PCR reactions. The primers were identical to those in Winton *et al*. (2007), but did not include the M13 universal tails resulting in shorter amplicons. A fluorescent dye label was included on each of the reverse primers to allow the amplified fragments to be distinguished from one another after genotyping. Fragment sizes were determined by capillary electrophoresis at the Oregon State University Center for Genome Research and Biocomputing (CGRB) Alleles were scored with the ABI GeneMapper 4.0 program. Lineage determination for each isolate was made according to the descriptions of multilocus genotypes (MLGs) described by Winton *et al*. (2007).

Data Analysis

The microsatellite genotypes were analyzed initially by using GenAlEx 6.5 (Peakall and Smouse 2006, 2012). Population genetics parameters obtained with this analysis included locus statistics, allelic diversity, and PhiPT estimates from an analysis of molecular variance (AMOVA). Further analyses were performed with R version 3.2.1 via R Studio. *Poppr 2.0.2* (Kamvar *et al*. 2014) was used to calculate genotypic diversity and to construct UPGMA dendrograms with bootstrapping. Adegenet 2.0.0 (Jombart 2008) was used for the discriminant analysis of principal components (DAPC) and K-means clustering.

Table 1. a) *Diversity statistics for each of the sampling sites.* ***b)*** *Diversity statistics for the two cryptic P. gaeumannii lineages. N= sample size, L1= abundance lineage 1, L2= abundance Lineage 2, MLG= number of multilocus genotypes, H= Shannon-Weiner diversity index.*

a)

Population	**N**	**L1**	**L2**	**MLG**	**H**
T-01	49	23	26	48	3.86
T-02	10	4	6	9	2.16
T-03	17	9	8	17	2.83
T5-3	29	22	7	29	3.37
T5-5	21	17	4	20	2.98
T15-1	46	23	23	45	3.80
T15-2	24	10	14	23	3.12
T25-2	99	99	0	90	4.43
T25-3	60	59	1	59	4.07
F15-1	29	27	2	27	3.27
F15-3	15	4	11	15	2.71
F25-2	37	25	12	32	3.41
DNRS25-1	21	13	8	25	3.22
DNRL25-2	25	19	6	16	2.71
14	**482**	**358**	**124**	**454**	**6.09**

b)

Lineage 1	358			335	5.78
Lineage 2	124			119	4.76
	482			**454**	**6.09**

RESULTS

Of the 482 isolates from 14 sites in Oregon and Washington included in this analysis, 454 (94%) possessed unique multilocus genotypes (MLGs) (Table 1). A total of 358 of these isolates had MLGs corresponding to previously reported alleles characteristic of Lineage 1, while 124 corresponded with Lineage 2 (Winton *et al*. 2007). Sites along the western Coast Range in northern Oregon and southern Washington, where SNC is most severe, consisted of roughly equal proportions of the two lineages and had relatively low genotypic diversity compared to isolates from sites further east. Diversity increased gradually in sites to the east of Tillamook, and was highest in the eastern Coast Range where Lineage 1 is more abundant (Figure 2, Figure 3). Only a single isolate of Lineage 2 was found at a site over 25 miles inland from Tillamook, Oregon (Table 1, Figure 3).

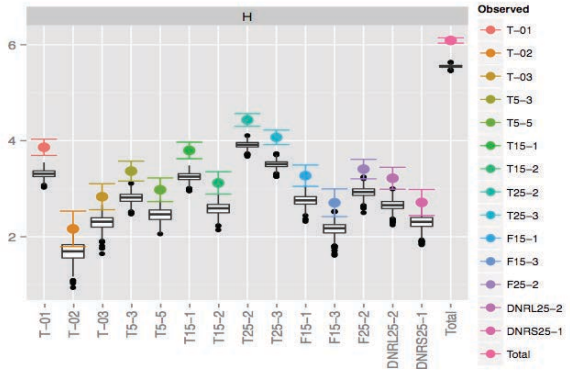


Figure 2. *Comparison of genotypic diversity by sampling site. H= Shannon-Weiner diversity index. Diversity was calculated on a rarefied sample of 10 isolates per site. Error bars represent 95% confidence intervals; boxplots represent values estimated from bootstrap replicates.*

A partitioning of molecular variance suggested that variation within subpopulations (sites or lineages) is greater than that between subpopulations. When sites were considered as subpopulations, within-site variation accounted for 90% of the total variation (Figure 4A), compared to 78% when lineages were considered subpopulations (Figure 4B). Estimates of subpopulation differentiation (PhiPT) were used to assess gene flow among populations. These analyses revealed that the populations across all sampled sites are only moderately differentiated from one another (PhiPT = 0.101, P = 0.001) (Figure 4A), whereas the lineages are strongly differentiated (PhiPT = 0.218, P = 0.001) (Figure 4B). The degree to which gene flow occurs between subpopulations was also assessed with a K-means clustering approach in which each isolate is assigned to a genetic group based on shared alleles, and the optimal number of genetic clusters in the data is determined based on estimates of the Bayesian Information Criterion (BIC). A total of 12 distinct genetic clusters were identified using this method. This technique was used in conjunction with the unweighted pair group method using arithmetic means (UPGMA) algorithm to analyze reproductive modes, population structure, and gene flow. The 12 clusters did not correspond to any distinct geographic group or lineage resulting in a mixing of genetic clusters across the branches of the UPGMA dendrogram (Figure 5A). This analysis was also performed with the isolates divided into 2 clusters without any *a priori* assignment to a lineage. This resulted in a clear assignment of the isolates to genetic clusters corresponding to the two lineages, as well as a correspondence of these clusters to the lineages on the UPGMA dendrogram (Figure 5B).

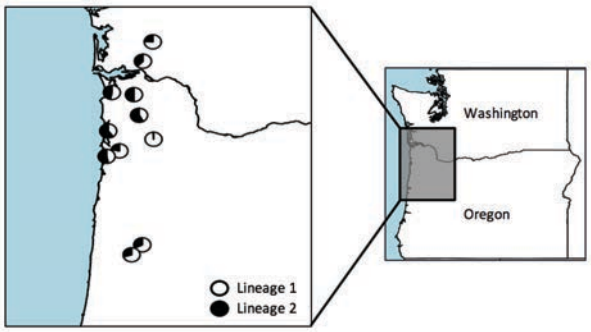
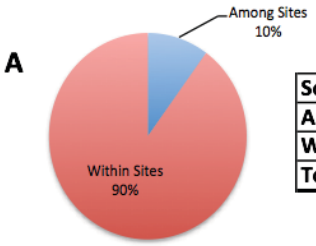
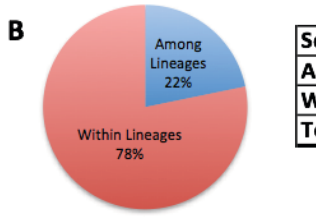


Figure 3. *Geographic distributions of the two P. gaeumannii lineages across the sampling sites. Isolates from sites T5-3 and T5-5, T25-2 and T25-3, F15-1 and F15-3 were pooled for this analysis to avoid overlap.*



Source	df	SS	MS	Est. Var.	%
Among Pops	13	228.713	17.593	0.416	10%
Within Pops	468	1743.715	3.726	3.726	90%
Total	481	1972.427		4.142	100%

Stat	Value	P(rand >= data)
PhiPT	0.101	0.001



Source	df	SS	MS	Est. Var.	%
Among Pops	1	171.691	171.691	1.044	22%
Within Pops	480	1800.736	3.752	3.752	78%
Total	481	1972.427		4.796	100%

Stat	Value	P(rand >= data)
PhiPT	0.218	0.001

Figure 4. *The results of an AMOVA comparing genetic variance within and among subpopulations.* ***(A)*** *Here, the subpopulations are the sampling sites.* ***(B)*** *Here, each lineage was considered a subpopulation. P-values from randomization test with 999 permutations.*

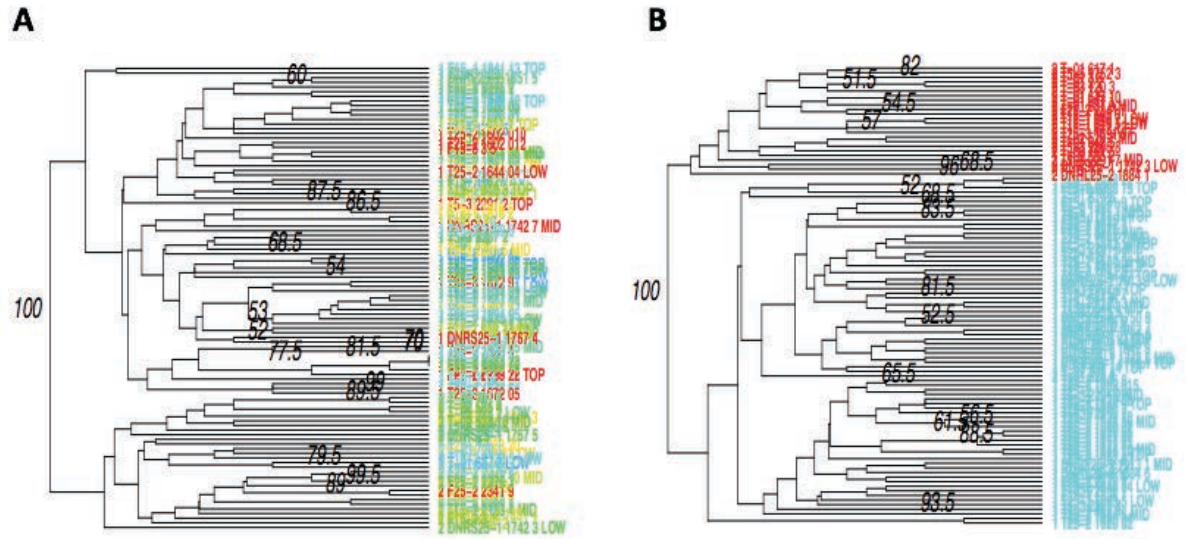


Figure 5. *UPGMA dendrograms constructed with 100 randomly selected isolates. The colors represent genetic clusters identified using a K-means approach. Node labels represent bootstrap statistics for 200 replicate trees.* ***A)*** *The optimum number of clusters (K=12) was chosen based on Bayesian information criterion (BIC).* ***B)*** *K=2 was chosen, and the two clusters correspond to the two lineages. Here, blue= Lineage 1, red= Lineage 2.*

The DAPC analysis was used to depict the relationship among subpopulations graphically. When sites were considered subpopulations, the isolates collected from T-02 and DNRS25-1 were significantly differentiated from those at all other sites, which formed a more compact cluster (Figure 6A). This differentiation was likely due to the presence of unique alleles present at only these two sites. When lineages were treated as subpopulations for this analysis, there was a distinct separation with no overlap, further supporting their strong differentiation and the lack of sexual recombination between them (Figure 6B).

DISCUSSION

Geographic centers of origin of plants, fungi, and microorganisms are generally characterized by high genetic diversity (Grünwald and Flier 2005). Populations of *P. gaeumannii* exhibit very high genotypic diversity and richness in the Pacific Northwest compared to regions where the fungus is known to be introduced, such as New Zealand (Bennett and Stone *this proceedings*), supporting the conclusion that this fungus is native to western North America. While Lineage 1 was found at every site sampled, and exhibited high diversity at sites in the eastern Coast Range, Lineage 2 was restricted to the western Coast Range and showed relatively low genotypic diversity. Lineage phylogeography seems to be correlated with the geographic distribution of SNC disease severity. In northwestern Oregon, the area where the most severe disease has been consistently found in annual aerial surveys corresponds to the area where the two lineages co-occur. There is a strong geographic trend of decreasing disease severity to the east of the Coast Range, where Lineage 2 is virtually absent. However, the significance of this correlation, and the factors influencing the restriction of the range of Lineage 2 to the western Coast Range are not fully understood.

If subpopulations of *P. gaeumannii* (site or lineage) are reproducing primarily by selfing (homothallism), or if gene flow is restricted due to geographic or genetic barriers, genetic clusters (groupings based on shared alleles) should be strongly associated with the geographic origins of the isolates. This would result in a sorting of isolates into geographic groups on the terminal branches of a UPGMA dendrogram, and a clear sorting of the genetic clusters into these geographic groups. Instead, our analyses do not support the interpretation that the 12 K-means clusters (Figure 5A) are related to geography or lineage. The K-means clustering algorithm identified multiple distinct clusters within each lineage from across their geographic ranges. This resulted in a mixing of isolates from different sites into the same K-means clusters. However when only two clusters were chosen for the K-means analysis, the resulting clusters corresponded exactly to the two lineages.

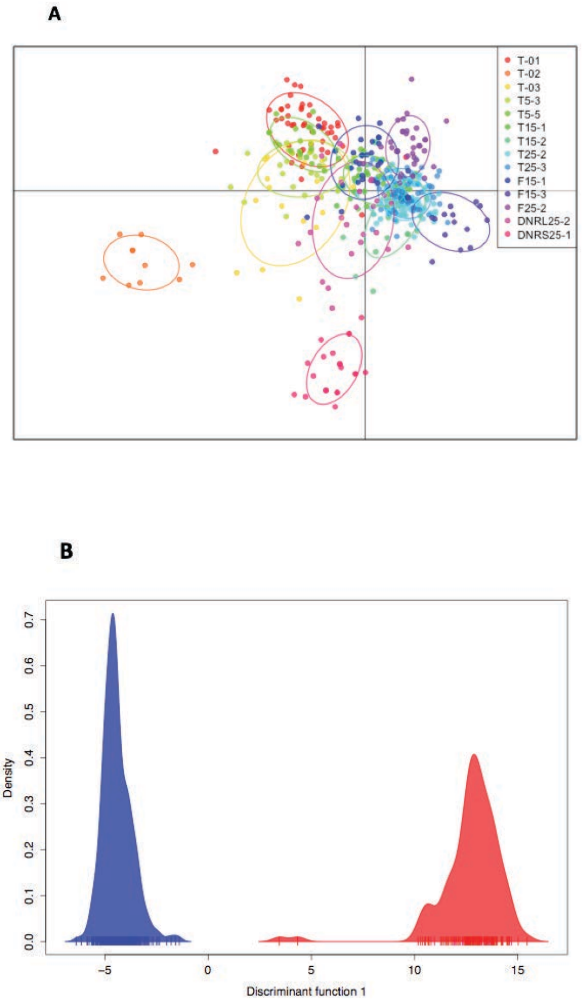


Figure 6. *DAPC plots showing relationships among subpopulations.* ***A****) Colors correspond with sampling sites, and points correspond to individuals.* ***B****) Colors correspond with lineages. Blue= Lineage 1, Red= Lineage 2.*

The separation of the two lineages at a basal node of the UPGMA dendrogram with high statistical support (Figure 5), and the sorting of isolates from the two lineages into distinct genetic clusters (Figure 5B) provides further evidence for the lack of recombination between lineages. These results are consistent with the assessments of differentiation determined with AMOVA.

The AMOVA showed that *P. gaeumannii* populations in the Pacific Northwest are heterogeneous in their genetic structure. Genetic differentiation is strong at the level of the lineage (PhiPT=0.218, P= 0.001), i.e.: the lineages are distinct subpopulations in our region, while the differentiation between sites is weaker (PhiPT= 0.101, P=0.001) suggesting the occurrence of dispersal and connectivity due to gene flow. The strong differentiation between the lineages is likely reflective of a form of reproductive isolation resulting from the predominance of homothallic reproduction and low rates of outcrossing, as suggested by Winton (2006). Although more variation occurred within lineages than between them, suggesting the existence of gene flow, the results of the UPGMA analyses, K-means clustering, and DAPC suggest otherwise. This AMOVA result is likely due to the existence of shared alleles at some of the loci that occurred through the process of convergent evolution rather than sexual recombination.

While the AMOVA allowed for the inference of gene flow and differentiation within and among subpopulations, the DAPC analysis allowed us to identify which, if any, subpopulations (sites or lineages) were significantly differentiated from the others. The results of this analysis, when sampling sites were considered subpopulations, suggested that two of the sites (T-02 and DNRS25-1) were differentiated from the other sites, and from one another (Figure 6A). While this may be due to the presence of rare or unique alleles at these sites, this may also be an artifact of homothallic reproduction resulting in an over-representation of genotypes from closely related individuals at these sites. This result could explain the moderate differentiation between sites revealed in the AMOVA analysis. The clear separation between isolates from the two lineages in the DAPC analysis (Figure 6B) provides further support for the lack of recombination between them, and also corroborates the results of the AMOVA analysis in which lineages were highly differentiated, suggesting a lack of gene flow between lineages.

The occurrence of homothallic reproduction, and the presence of significant subpopulation structure in populations of *P. gaeumannii* may have implications for the evolutionary adaptive potential of *P. gaeumannii*, and thus may have an impact on SNC management strategies and the mitigation of growth impacts in commercial timber production. In addition to the climate variables known to influence SNC disease distribution and severity, the relative proportions of the two lineages present at a site should be considered when making an assessment of future SNC disease risk.

REFERENCES

Boyce, J.S. 1940. A needle-cast of Douglas Fir associated with *Adelopus gaeumannii*. Phytopathology 30:649-659.

Grünwald, N.J. and Flier W.G. 2005. The biology of *Phytophthora infestans* at its center of origin. Annual Reviews in Phytopathology 43:171–190.

Jombart, T. 2008. Adegenet: an R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Kamvar Z.N., Tabima J.F., and Grünwald N.J. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281.

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

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

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ritokova, G. *et al*. 2013. SNCC Research and Monitoring Plot Network. Swiss Needle Cast Cooperative Annual Report. College of Forestry, Oregon State University, Corvallis, Oregon.

Winton, L.M., Hansen, E.M., and Stone, J.K. 2006. Population structure suggests reproductively isolated lineages of *Phaeocryptopus gaeumannii*. Mycologia 98:781–791.

Winton, L.M., Stone, J.K. & Hansen, E.M. 2007. Polymorphic microsatellite markers for the Douglas-fir pathogen *Phaeocryptopus gaeumannii*, causal agent of Swiss Needle Cast disease. Molecular Ecology Notes 7**:**1125–1128.