

Supplementary Materials: Potential Population Genetic Consequences of Habitat Fragmentation in Central European Forest Trees and Associated Understorey Species—An Introductory Survey

Christoph Dobeš, Heino Konrad and Thomas Geburek

Supplementary Material 2 Statistic analyses—Analyses were based on the data provided in Supplementary Material 1 which represent 53 Central European tree species for which records on population genetic studies providing estimates of population genetic differentiation are available or allowed for its calculation from the original data. Five records were excluded from analyses because some populations were either strongly introgressed (one record for *Quercus robur* and two records for *Q. petraea* [1,2]), showed pronounced clonal population structure (*Sorbus domestica* [3]), or because differentiation was calculated among regions only (*Fagus sylvatica* [4]). For analyses the data were organized into three subsets: records on biparentally—(i.e., nuclear, final sample size = 198), maternally—(chloroplast markers in angiosperms and mitochondrial markers, $N = 85$), and paternally-inherited markers (chloroplast markers in gymnosperms, $N = 15$). Statistics performed involved application of unifactorial ANOVAs for which we used *Statistica* v12 [5], Generalized Linear Models (GLM) and linear regressions both performed using the programming environment R [6], and descriptive statistics for which both R and MicroSoft Excel were used.

Preselection of Marker Systems and of Measures of Population Differentiation

As a methodological prerequisite we examined in a further step marker systems and types of differentiation measures for possible mutual deviation in their estimates of population differentiation.

Biparentally-inherited markers: Seven marker systems (AFLPs, isozymes, ISSRs, RAPDs, RFLPs, SNPs, and SSRs) and eight types of differentiation measures (d_0 , δ , D_j , F_{st} [including θ], G_{st} , Φ , R_{st} , Shannon-Weaver) have been used by the various authors in studies based on biparentally-inherited markers. As for marker systems we reduced comparisons to the codominant systems and AFLPs since the remaining dominant markers were highly underrepresented in the data ($N = 1$ –4 records). Estimates obtained using isozymes and SSR did not differ significantly from each other but both differed significantly from those based on SNPs as well as AFLPs (unifactorial ANOVA; Table S1). We limited further analyses to the combined records on isozymes and SSRs (219 out of 252 biparentally-based records). F_{st} , G_{st} , and Φ were the most frequently used measures in the reduced dataset (198 records). These measures are known to give in most cases similar values [7]. This assumption has been tested using unifactorial ANOVA and found valid (Table S2), a result also supported by the GLM (Table S3). In order to reduce the complexity of the data we excluded the other measures present at low frequencies and for which we could not perform meaningful statistic comparisons.

Table S1. Comparison of estimates of genetic differentiation of tree populations for biparentally-inherited markers. unifactorial ANOVA, Bonferroni post-hoc test; df = 220; significance codes: *** 0.001; ** 0.01; * 0.05.

	Isozymes	SNPs	SSRs
SNPs	**		
SSRs		***	
AFLPs	**	***	*

Table S2. Comparison of types of measures of genetic differentiation for biparentally-inherited codominant markers (isozymes and SSRs). unifactorial ANOVA, Bonferroni post-hoc test; df = 195; significance codes: none of the comparisons was significant at $p \leq 0.05$.

	F_{st}	G_{st}
G_{st}	n.s.	
Phi	n.s.	n.s.

Table S3. GLM for biparentally-inherited markers: full model and reduced model. Models were run for all parameters and with one of each of the parameter excluded.

	df	Deviance	AIC	Scaled Dev.	p (>Chi)	
full model						
all parameters		110.77	445.35			
mode of diaspore dispersal	1	111.89	445.13	1.784	0.1817	
mode of pollen dispersal	2	122.86	459.68	18.334	0.0001	***
marker system (isozymes and SSRs only)	1	111.33	444.25	0.897	0.3436	
type of differentiation measure	3	111.16	439.97	0.625	0.8907	
N populations	1	110.78	443.36	0.014	0.9052	
mean population size	1	123.44	462.51	19.163	1.200e-05	***
N loci	1	123.51	462.61	19.262	1.139e-05	***
spatial extension study area	1	139.02	483.55	40.201	2.291e-10	***
reduced model						
all parameters		112.90	436.71			
mode of pollen dispersal	2	125.06	450.82	18.112	0.0001	***
mean population size	1	124.74	452.37	17.662	2.639e-05	***
N loci	1	127.01	455.56	20.847	4.975e-06	***
spatial extension study area	1	142.37	475.76	41.050	1.484e-10	***

Significance codes: *** 0.001

Uniparentally-inherited markers: Records of maternally-inherited markers involved save for a single exception (for which R_{st} is provided) estimates based on F_{st} , G_{st} , and Φ only (85 records) and only these were kept for further analyses. Four marker systems were applied in the underlying studies (mini-satellites, RFLPs, SNPs, and SSRs). We evaluated effects of type of marker system on the estimate of population genetic differentiation in the GLM. Dropping this variable from the model did not deteriorate the AIC suggesting that the different marker systems had no significant effect and consequently were kept in the analyses (Table S4). Estimates for paternally-inherited markers again were mainly based on F_{st} , G_{st} , and Φ (17 records; for two records R_{st} was used) and data were reduced to these records. In one study each mini-satellites and SNPs, respectively, served as marker system, which were excluded, and only remaining 15 SSR-based records were kept.

Predictive Value of Tree Life History Traits and of Characteristics of Study Design on Population Differentiation

GLMs were used to evaluate the value of life history traits of trees (mode of diaspore dispersal, mode of pollen dispersal, and mode of inheritance) and of characteristics pertaining to study design (number of studied populations, mean number of individuals analysed per population, number of screened loci, and spatial extension of the study area) in predicting population differentiation. Two life history traits presented in Supplementary Material 1 were excluded for the following reasons: Breeding system because data on mating system were too incomplete and we felt that thorough description of its expression in the treated tree species would have needed a separate review which was out of scope of this survey. Distribution of sexes was further non-random with respect to mode of pollen dispersal (i.e., monoecy was associated with wind-pollination only) potentially leading to confounding effects in the GLM. For the same reason we didn't include ploidy in the analyses since mixed ploidy was associated with endozoochory and hermaphroditism only.

GLMs were performed using the *glm* function implemented in the *stats* package. The dependent variable *population differentiation* was first transformed to obtain a normal distribution of values using the *BoxCox* function of the *hdcde* package. As options of the GLM we choose *Gaussian* as family and *identity* as link function. We started with a full model which included all variables (i.e., predictors) and then applied the *drop1* function (which runs the same model but with one variable excluded) to evaluate the relative importance of the variables. A significant increase of the AIC value upon exclusion of a variable compared to the AIC of the full model signify that the variable is relevant in predicting population differentiation. Subsequently a reduced model with all predictors found significant was run to which the *drop1* function was again applied. GLMs were applied to records using biparentally—(Tables S3) and maternally-inherited markers (Tables S4), but not to those based on paternally-inherited markers due to low numbers of records available for the latter. Characteristics of significant predictive value—but all life history traits—were subsequently explored using unifactorial ANOVAs and the Bonferroni post-hoc test for pairwise comparison of categories (Tables 1 and 2 in the main text) for categorical variables and linear regressions for continuous variables (Figure 4 main text).

Table S4. GLM for maternally-inherited markers: full model. The model was run for all parameters and with one of each of the parameter excluded.

	df	Deviance	AIC	Scaled Dev.	<i>p</i> (>Chi)	
full model						
all parameters		254401	862.49			
mode of diaspore dispersal	1	255648	860.86	0.3716	0.542133	
marker system	5	266495	856.02	3.5298	0.618881	
type of differentiation measure	3	275849	862.64	6.1517	0.104459	
N populations	1	292791	871.17	10.6819	0.001082	**
mean population size	1	255541	860.83	0.3399	0.559882	
N loci	1	292644	871.13	10.6434	0.001105	**
spatial extension study area	1	257980	861.55	1.0619	0.302781	
reduced model						
all parameters		322700	939.69			
N populations	1	366332	948.34	10.6525	0.001099	**
N loci	1	360407	946.97	9.2828	0.002313	**

Significance codes: ** 0.01

References

- Samuel, R.; Pinsker, W.; Ehrendorfer, F. Electrophoretic analysis of genetic variation within and between populations of *Quercus cerris*, *Q. pubescens*, *Q. petraea* and *Q. robur* (Fagaceae) from Eastern Austria. *Bot. Acta* **1995**, *108*, 290–299.
- Kelleher, C.T.; Hodkinson, T.R.; Kelly, D.L.; Douglas, G.C. Characterisation of chloroplast DNA haplotypes to reveal the provenance and genetic structure of oaks in Ireland. *For. Ecol. Manag.* **2004**, *189*, 123–131.
- George, J.P.; Woodman, J.; Hampton, M.; Konrad, H.; Geburek, T. True Service-tree (*Sorbus domestica*, Rosaceae) in the British Isles: Rare but diverse. *New J. Bot.* **2016**, *6*, 21–30.
- Paule, L.; Gömöry, D.; Vysny, J. Genetic diversity and differentiation of beech populations in Eastern Europe. Genetics and silviculture of beech. In Proceedings from the 5th Beech Symposium of the IUFRO Project Group P1.10-10, Mogenstrup, Denmark, 19–24 September 1994; Madsen, S., Ed.; Danish Forest and Landscape Research Institute: Hørsholm, Denmark, 1995; pp. 159–167.
- StatSoft, I. STATISTICA für Windows. Software System für Datenanalyse. Version 12. 2013. Available online: www.statsoft.com (accessed on 10 January 2016).
- R-Development-Core-Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2011. Available online: <http://www.R-project.org/> (accessed on 9 January 2016).

7. Meirmans, P.G.; Hedrick, P.W. Assessing population structure: F_{st} and related measures. *Mol. Ecol. Res.* **2011**, *11*, 5–18.



© 2017 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).