

# Bayesian approach reveals confounding effects of population size and seasonality on outcrossing rates in a fragmented subalpine conifer

Igor J. Chybicki · Artur Dzialuk

Received: 27 January 2014 / Revised: 30 June 2014 / Accepted: 19 August 2014 / Published online: 31 August 2014  
© The Author(s) 2014. This article is published with open access at Springerlink.com

**Abstract** Mating systems have long been recognized as key factors determining genetic structure within and between populations. Outcrossing promotes genetic diversity and gene flow between populations, while inbreeding, on the other hand, decreases recombination rates, facilitating fixation of co-adapted genes. In small populations, selfing moderates pollen limitation because of low mate availability, but at the cost of increased inbreeding depression. These conflicts are of more than theoretical interest; they are critical for the management of endangered species. In order to help designing conservation strategies for the management of the gene pool of fragmented populations of *Pinus cembra*, a protected species in Poland, we have characterized pollen flow and mating structure using nuclear microsatellite markers. We demonstrated that *P. cembra* in the studied stands of the Tatra Mts. is characterized by an average outcrossing rate ( $t$ ) of 0.72. Unlike with the existing approaches, using the newly developed Bayesian method, we found that population size and seasonal variation had confounding effects on outcrossing rates. In concordance with predictions, large populations showed significantly higher outcrossing rates ( $t=0.89$ ) than smaller ones ( $t=0.51$ ). Temporal variation revealed in the outcrossing rate might be linked with masting behavior of the species. On the other hand, we showed that outcrossing rates were not associated with a trunk diameter of a mother tree. Our study also demonstrated that biparental inbreeding is a significant component of mating system. However, we further show that pollen dispersal follows a fat-tailed distribution (with the average dispersal distance of 1,267 m) so that at least some long-distance pollen dispersal must be occurring.

Overall, we conclude that the high inbreeding (both selfing and mating between relatives) found in *P. cembra* buffers for pollen limitation. We argue that small, isolated stands can be at risk of gene pool erosion, despite the potential for long-distance pollen and seed dispersal.

**Keywords** Population fragmentation · Mating system · Biparental inbreeding · Pollen dispersal · Bayesian method · *Pinus cembra*

## Introduction

Mating system has long been recognized as a key factor determining genetic structure within and between populations (Wright 1931; Jain and Marshall 1985; Mitton 1992). In plants, two mating systems are generally recognized, i.e., self- and cross-fertilization. Outcrossing promotes genetic diversity and gene flow between populations (Mitton 1992). Moreover, outcrossing is advantageous because it allows individuals to avoid homozygosity for deleterious mutations, which leads to inbreeding depression (Lande and Schemske 1985; Charlesworth and Charlesworth 1987). However, inbreeding depression may not be the main factor affecting mating system. The survey made by Goodwillie et al. (2005) showed that many species under strong inbreeding depression maintain high selfing rates. On the contrary, some species favor outcrossing even if inbreeding depression is low. On the other hand, self-fertilization can be advantageous because it results in 50 % gain of gene transmission as compared with outcrossing. Self-fertilization increases individual homozygosity and consequently decreases recombination rates (Jain 1976; Wright et al. 2008). Hence, self-fertilization can be an efficient mechanism that enhances fixation of co-adopted gene complexes (Lande and Schemske 1985). All these forces shape the distribution of selfing rates in natural populations. Interestingly, this distribution is more or less continuous in

Communicated by G. G. Vendramin

I. J. Chybicki (✉) · A. Dzialuk  
Department of Genetics, Kazimierz Wielki University,  
Chodkiewicza 30, 85064 Bydgoszcz, Poland  
e-mail: igorchy@ukw.edu.pl

animal-pollinated species, whereas in wind-pollinated species, it is clearly bimodal, with both extremes being over-represented as compared with the total (i.e., irrespective of pollination mode) distribution (Vogler and Kalisz 2001 and references therein). Thus, it seems that extreme mating systems prevail in the case of wind pollination.

Among other factors influencing mating systems, self-incompatibility and pollen availability need particular attention (Young et al. 2012). In many plants (e.g., Brassicaceae, Rosaceae, or Solanaceae), genetically determined self-incompatibility is an efficient way to recognize and reject self-pollen, preventing not only self-fertilization but also biparental inbreeding (reviewed in Takayama and Isoagi 2005). The second factor is related to the hypothesis of reproductive assurance. Predominantly outcrossing species may experience lower fitness when pollen sources or pollinators are scarce. In this case, selfing can be selected as a means to assure reproduction, even at the cost of increased inbreeding depression (Lloyd 1979; Holsinger 1996; Morgan and Wilson 2005; Porcher and Lande 2005). Also, the mass action model predicts that selfing is both density- and frequency-dependent because the rate of self-fertilization increases as the rate of self-pollination (i.e., the relative amount of self- to total pollen on receptive stigmas) increases (Holsinger 1991). For this reason, many species display the mixed mating system, combining the advantages of both extreme reproductive strategies (Goodwillie et al. 2005).

The factors that influence mating system can shape variation in outcrossing rates between populations and seasons or even among individuals. Such a variation can arise, for example, from a variable amount of pollen being available within different populations or seasons. It has been often demonstrated that population density can have impact on selfing rates (Farris and Mitton 1984; Knowles et al. 1987; El-Kassaby and Jaquish 1996; Robledo-Arnuncio et al. 2004), including the break-down of the self-incompatibility system (Kamm et al. 2011). Generally, variation in self-fertilization can lead to unpredictable oscillations in inbreeding rates (Coelho and Vencovsky 2003), and thus, may be a basis for stronger inbreeding effects than predicted by simple models. In small populations, variation in outcrossing rates ( $t$ ) can result in fluctuation of  $t$  between generations (Jain and Marshall 1985), leading to high variation in fixation rates.

Here, we investigated the impact of the size of isolated population fragments on outcrossing rate. As a case study, we used Swiss stone pine populations located in Tatra Mts., Poland. Swiss stone pine (*Pinus cembra* L.) is a European coniferous tree occupying the timberline ecotone in the Alps and the Carpathians. Unlike to the Alpine core range, where *P. cembra* is locally abundant, small isolates in Carpathians are relatively common. At the end of the nineteenth century, the main priority of stone pine restoration in Poland was the

genetic rescue by sowing and planting stone pines of local but also Alpine or even Siberian origin (Paryski 1971). Since 1946, stone pine is protected by Polish law, and since 1954, when the Tatra National Park was established, the whole Polish population of the species was included within the nature reserve (Zwijacz-Kozica and Żywiec 2007). Unfortunately, stone pine is still believed to be a threatened species in the Polish Tatras (Zwijacz-Kozica and Żywiec 2007). Due to negative anthropogenic impact, most of its stands have been decimated, and as a consequence, stone pine occurs today mainly in almost inaccessible areas (Zwijacz-Kozica and Żywiec 2007), on poor, non-calcareous bedrock, mainly on granite (Myczkowski and Bednarz 1974). At present, there are about 12,000 individuals of stone pine in the entire Tatra region (Jamnický 1981). However, only 30 % of the total is located in the Polish Tatras, growing mostly as isolated groups (Chmiel 1996).

To date, the mating system of *P. cembra* in the Tatra Mts. has not been characterized. However, the results for Alpine and East Carpathian populations revealed that *P. cembra* is characterized by slightly lower outcrossing rates as compared with other pines (Lewandowski and Burezyk 2000; Politov et al. 2008; Salzer and Gugerli 2012; cf. Restoux et al. 2008), with the average about 75 %. However, outcrossing rate can be even lower than estimated based on genotyped seeds (or seedlings). This is because early-acting inbreeding depression can lead to abortion of selfed embryos (Goodwillie et al. 2005). In fact, no correlation was found between fitness components and selfing rates measured at the seed stage in *P. cembra* (Salzer and Gugerli 2012). However, the results indicated that the rate of early abortion of embryos due to inbreeding depression was higher in small peripheral populations compared to more continuous ones. Interestingly, a detailed study in the Alps showed that population fragmentation significantly reduces outcrossing (Salzer and Gugerli 2012). This can occur as a result of pollen limitation. However, pollen limitation requires pollen production and/or dispersal to be significantly reduced in comparison with a large, more continuous population.

In this study, we aimed to verify the hypotheses that (a) outcrossing rates were positively associated with population size and (b) outcrossing rates revealed seasonal invariance. For this purpose, a new method was developed, which allows the accounting for multiple confounding factors of outcrossing rates, measured either in a continuous or a nominal scale. Furthermore, we focused on verifying whether mating between relatives contributed to the total rate of inbreeding. Finally, we aimed at assessing pollen dispersal kernel with a special emphasis on whether pollen gametes have the potential for long distance dispersal.

## Materials and methods

### Study species

The Swiss stone pine (*P. cembra*) is a monoecious five-needled pine, growing at the timberline in the European mountains. The gene flow can theoretically be expected to be extensive in Swiss stone pine, since the species is a predominantly outcrossed, with wind-dispersed pollen, and its wingless seeds are almost exclusively dispersed by the far-ranging bird European nutcracker (*Nucifraga caryocatactes*; Ulber et al. 2004). The spatial genetic structure within populations of *P. cembra* can show aggregation of genetically related individuals as a result of caching behavior of the nutcracker. Kin-structured regeneration can potentially fosters mating among relatives (Salzer 2011). Moreover, flowering and seed production in Swiss stone pine occurs every 2–3 years, but only a single year of 4–10 is an abundant mast year (Ulber et al. 2004). In consequence, temporal fluctuations in mating patterns are possible.

### Study material

Because *P. cembra* in the Tatras occupies timberline (typically at 1,300–1,700 m a.s.l.), subpopulations' borders can be easily delineated through the system of mountain valleys. Consequently, subpopulations are represented by geographically isolated stands. We sought to collect seeds within 4 stands for 3 successive seasons (Fig. 1). Among selected localities, Dolina Białego and Dolina Suche Kasprowej represent relatively small populations, with about 50 and 350 adult individuals, respectively. The other two, Morskie Oko and Dolina Waksmundzka, represent relatively large populations, with about 1,500 adult individuals each (Myczkowski and Bednarz 1974). The mean geographic distance between stands was 7.3 km. Due to hazard sampling conditions (a very steep slope or cliffs), within each locality, only five mother trees were initially selected. The selected trees were kept the same over the three sampling seasons. In order to avoid additional sources of variation, we did not sample nearest neighbors or trees close to the subpopulation border. The average distance between trees within a subpopulation ranged from 28 (Morskie Oko) to 109 m (Dolina Białego), with the average of 61 m.

In June, four metal net bags of 40 cm×50 cm were put on branches with young cones on each mother tree to protect seeds against the harvesting by the mountain jay (*N. caryocatactes* L.). Two-year-old seeds were sampled in October. However, due to low seed productivity as well as incidental loss of the material (destruction of bags) or tree fall (a single case), only 13 individuals (mother trees) were successfully sampled. In total, 1,929 seeds were collected (522 in 2008, 1,167 in 2009, and 240 in 2010). In order to have as

much as possible balanced seed samples, the random subsample of 1,129 seeds was only analyzed. The number of seeds used for genotyping ranged from 31 to 131 with the average of 87 per mother tree (Table 1). From each seed, the embryo was extracted and used for isolation of the total genomic DNA (using CTAB protocol; Doyle and Doyle 1990).

### Genotyping

Embryos extracted from seeds were genotyped at 7 nuclear microsatellite markers (Pc1b, Pc3, Pc7, Pc18, Pc22, Pc23, and Pc35) developed for *P. cembra* (Salzer et al. 2009) using the Multiplex Master Mix (Qiagen) according to the manufacturer's instruction. In order to describe the overall variation of genetic markers, the following genetic parameters were computed: the number of alleles (*A*), the observed (*H<sub>o</sub>*), and expected (*H<sub>e</sub>*) heterozygosity. Additionally, in order to describe the power of genetic markers, we also estimated the exclusion probability (Eq. 1 in Jamieson and Taylor 1997).

### Analysis of outcrossing rates

Outcrossing rates were estimated using several approaches. Due to the sampling design, we were able to organize seeds into groups according to a population size and a year of seed collection. Therefore, besides the overall average, outcrossing rates were estimated separately per population size and per year. Both multilocus (*t<sub>m</sub>*) and single-locus estimates (*t<sub>s</sub>*) were obtained using the maximum likelihood approach implemented in MLTR 3.4 (Ritland 2002). The single-locus approach is useful to quantify the overall level of inbreeding (i.e., both due to selfing and mating between relatives), while *t<sub>m</sub>* is strictly related with selfing. Therefore, the difference *t<sub>m</sub>*–*t<sub>s</sub>* can be used as to measure the level of biparental inbreeding (Ritland 2002). Correlation of selfing rate among loci, *r<sub>s</sub>*, provides additional information about biparental inbreeding. If there is no biparental inbreeding then *r<sub>s</sub>*=1. Additionally, (1–*r<sub>s</sub>*) is approximately the fraction of inbreeding due to biparental inbreeding. Standard errors of the estimates were computed using the bootstrap procedure (with 1,000 bootstrap samples).

### Effects of population size and seasonal variation and individual outcrossing probabilities

Simple statistical comparisons of mean outcrossing rates between seasons and between large and small populations accounted for a single factor at once. However, in the case of confounding effects of the two factors, such comparisons can be misleading. Knowing exactly whether a given seed has been a result of outcross fertilization vs. self-fertilization, one could use the multiple logistic regression analysis with nominal explanatory variables (i.e., *season* and *population size*; Table 1). In our case, however, such information was not



**Fig. 1** Location of populations of *Pinus cembra* in Tatra Mts. and the species natural range in Europe (gray-shaded areas). Distribution map courtesy to EUFORGEN, [http://www.euforgen.org/distribution\\_maps.html](http://www.euforgen.org/distribution_maps.html)

available a priori. To cope with this, we developed a mixed mating model-based regression analysis, in which unknown outcrossing rates were expressed as a function of the explanatory variables.

Generally, in order to perform the regression analysis, for each nominal variable  $K-1$  unredundant code (or design) variables need to be created, where  $K$  is a number of levels observed at a given variable (Hosmer and Lemeshow 2000).

**Table 1** Number of seeds analyzed in particular years and populations

Family	Population	UTM coordinates <sup>a</sup>	DBH <sup>b</sup>	2008	2009	2010	Total	Outcrossed (%)
Small populations								
1	Dolina Białego	423901,5456781	40	-	48	7	55	15 (27.3)
2	Dolina Suchej Kasprowej	425849,5454810	41	31	-	-	31	17 (54.8)
3	Dolina Suchej Kasprowej	425875,5454770	48	45	34	30	109	57 (52.3)
4	Dolina Suchej Kasprowej	425897,5454743	61	37	-	15	52	43 (82.7)
5	Dolina Suchej Kasprowej	425899,5454806	50	40	46	21	107	23 (21.5)
6	Dolina Suchej Kasprowej	425908,5454840	44	40	46	45	131	92 (70.2)
Large populations								
7	Dolina Waksmundzka	432603,5455253	50	36	47	9	92	74 (80.4)
8	Dolina Waksmundzka	432625,5455259	53	36	49	45	130	125 (96.2)
9	Morskie Oko	432515,5449729	60	35	-	-	35	29 (82.9)
10	Morskie Oko	432537,5449744	63	42	45	-	87	75 (86.2)
11	Morskie Oko	432536,5449729	NA <sup>c</sup>	50	43	-	93	77 (82.8)
12	Morskie Oko	432558,5449724	35	42	46	28	116	105 (90.5)
13	Morskie Oko	432535,5449720	38	49	42	-	91	80 (87.9)
Total				483	446	200	1,129	812 (71.9)

Dashes indicate years when seeds were not available. The column “Outcrossed” shows the number of seeds being a result of outcrossing (inferred based on the Bayesian paternity analysis; for details, see the “Materials and methods” section)

<sup>a</sup> Geographic position of a mother tree

<sup>b</sup> Diameter at breast height of a mother tree (in cm)

<sup>c</sup> Not available

To create codes, the method of “deviations from means” (also known as the “effects coding”) was chosen (Hosmer and Lemeshow 2000) because in this case regression slopes are informative about a direction and strength of levels’ effects relative to the mean effect of a nominal variable. It should be stressed, however, that the overall statistical behavior (e.g., the overall fit) of the model remains invariant in respect to the choice of a coding system. In our case, the variable *season* had three levels (2008, 2009, 2010), while the variable *population size* had two levels (small, large; Table 1). Therefore, in total, four code variables were created (Table 2).

In addition, we included a diameter at breast height (DBH) as a potential factor of outcrossing rate. Here, DBH was used as a proxy for male fecundity (an amount of pollen produced by a tree). According to the mass action model (Holsinger 1991), where the self-fertilization rate is proportional to the amount of self pollen relative to the total pollen, male fecundity may have a confounding, and presumably, negative effect (see, e.g., Shea 1987 for empirical results) on the expected outcrossing proportions, affecting potentially the results for small and large populations.

Finally, using the logistic regression model, the outcrossing rate (a response variable) for the  $m$ -th ( $m=1,2,3$ ) *season*, the  $n$ -th ( $n=1,2$ ) *population size*, and the maternal DBH equal to  $d$  were formulated as follows:

$$t_{mnd} = \frac{\exp(\alpha + \beta_{sm} + \beta_{pn} + \gamma d)}{1 + \exp(\alpha + \beta_{sm} + \beta_{pn} + \gamma d)} \quad (1)$$

where  $\alpha$  is the intercept,  $\beta_{sm}$  is the regression slope coefficient for the effect of the  $m$ -th level in the *season* variable,  $\beta_{pn}$  is the regression slope coefficient for the effect of the  $n$ -th level in *population size* variable,  $\gamma$  is the regression slope coefficient for the effect of DBH, and  $d$  is the DBH of a mother tree. Note that in the Eq. (1),  $\beta_{s3} = -(\beta_{s1} + \beta_{s2})$  and  $\beta_{p2} = -\beta_{p1}$  (see Table 2), so that the number of unconstrained (estimable)  $\beta$  coefficients remains equal to 4 (the same as the number of design variables).

To estimate the regression parameters, we adopted the Bayesian approach developed recently as a framework to model plant mating systems (Chybicki 2013; Chybicki and Burczyk 2013). The approach is generally based on the (mixed mating) probability model, similar to that used in MLTR, in which the probability of the multilocus genotype  $O_{ij}$  of the  $j$ -th offspring in the  $i$ -th maternal family is as follows:

$$\Pr(O_{ij}|m = Season_{ij}, n = Popsiz_{e_i}, d = DBH_i) = (1 - t_{mnd})\Pr(O_{ij}|M_i) + t_{mnd}\Pr(O_{ij}|M_i, \mathbf{P}) \quad (2)$$

where  $t_{mnd}$  is the probability that a seed collected in the  $m$ -th year and the  $n$ -th population type taken from a mother tree of DBH equal to  $d$  was produced through outcrossing,  $\Pr(O_{ij}|M_i)$  is the Mendelian probability of the offspring genotype after self-fertilization given the maternal genotype  $M_i$ ,  $\Pr(O_{ij}|M_i, \mathbf{P})$  is the Mendelian probability of the offspring genotype after outcrossing given the maternal genotype  $M_i$  and the background pollen pool  $\mathbf{P}$  represented by the matrix of allele frequencies, while  $Season_{ij}$  is the collection year of the  $j$ -th seed in the  $i$ -th family,  $Popsiz_{e_i}$  is the level of the *Population size* into which the  $i$ -th mother tree is classified (see Table 2), and  $DBH_i$  is the diameter at breast height of the  $i$ -th mother tree. In the case of the tree 11, the missing value was replaced with the arithmetic mean (49 cm). In the Eq. (2),  $t_{mnd}$  was substituted with the right-hand side of the Eq. (1), so that the likelihood of the model

$$L(\mathbf{O}; \alpha, \beta, \mathbf{P}) = \prod_i \prod_j \Pr(O_{ij}|m = Season_{ij}, n = Popsiz_{e_i}, d = DBH_i) \quad (3)$$

was actually a function of regression coefficients and not outcrossing rates (as in MLTR).

The parameters were estimated using the Gibbs sampler (a class of Markov Chain Monte Carlo or MCMC algorithm). In this approach, the posterior distribution is approximated by a sequence of parameter values cyclically drawn using the specified sampling scheme. A uniform Dirichlet distribution was taken as a prior for  $\mathbf{P}_l$  (i.e., the vector of background allele frequencies at the  $l$ -th locus), while for the regression parameters  $\alpha$ ,  $\beta$ , and  $\gamma$ , a zero-centered normal distribution was taken with the variance of 10. For technical reasons (cf. Chybicki 2013; Chybicki and Burczyk 2013), the additional auxiliary variables  $\mathbf{A}$  and  $\mathbf{X}$  were introduced. Each element  $A_{lk}$  in the matrix  $\mathbf{A}$  was used to store the number of copies of the  $k$ -th allele at the  $l$ -th locus in a background pollen pool. Note that  $\mathbf{A}$  stores counts of alleles present in outcross pollen gametes only (Chybicki 2013). The elements  $X_{ij}$  in  $\mathbf{X}$  equaled 1 or 0 if the  $j$ -th seed in the  $i$ -th family was produced through outcrossing or selfing, respectively. The auxiliary variables are normally unknown, but they can be inferred from data given  $\mathbf{t}$  and  $\mathbf{P}$ , using a Bernoulli scheme described in detail in Chybicki (2013) and Chybicki and Burczyk (2013). Here, we only stress that, in a consequence of the algorithm,  $\mathbf{P}$  was estimated simultaneously with the remaining parameters (as in MLTR software). The vector  $\{A_{lk}\}$  has a multinomial distribution with the parameter  $\mathbf{P}_l$ . Thus, because a Dirichlet distribution is a conjugate prior for a multinomial distribution,  $\mathbf{P}_m$  could be sampled directly from the posterior multinomial-Dirichlet distribution, leading to the efficient Gibbs sampler for allele frequencies. Summing up, a single MCMC cycle started with inferring  $\mathbf{A}$  and  $\mathbf{X}$ . Then, for every  $l$ -th locus,  $\mathbf{P}_l$

**Table 2** Specification of the code (design) variables for the nominal predictors (*season* and *population size*) and the corresponding regression function terms

Variable	Level	Code (design) variables		Regression function term
<i>season</i>	1 (2008)	$x_{s1}$ 1	$x_{s2}$ 0	$\beta_{s1}x_{s1} + \beta_{s2}x_{s2} = \beta_{s1}$
	2 (2009)	0	1	$\beta_{s1}x_{s1} + \beta_{s2}x_{s2} = \beta_{s2}$
	3 (2010)	-1	-1	$\beta_{s1}x_{s1} + \beta_{s2}x_{s2} = -(\beta_{s1} + \beta_{s2})$
<i>population size</i>	1 (small)	$x_{p1}$ 1		$\beta_{p1}x_{p1} = \beta_{p1}$
	2 (large)	-1		$\beta_{p1}x_{p1} = -\beta_{p1}$

was drawn using the Gibbs sampler. Then, the parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  were updated using the Metropolis–Hastings algorithm, in which new values were proposed using the reflective random walk (Hoff 2009). The final estimates (posterior means and quantiles) were obtained after 1,000,000 cycles, keeping every 250th update and disregarding the first 100,000 updates for burn-in. A given nominal variable was considered as statistically significant, if any of the corresponding slope coefficients estimated for this variable was significantly different from zero (based on the credible interval). As a cross-validation of this procedure, we performed the Bayesian model comparison based on the DIC (Deviance Information Criterion, Spiegelhalter et al. 2002). For this purpose, additional regression analyses were performed based on the simplified (nested) models, resulting from removing one, i.e., *season* or *population size*, or both variables (the null model).

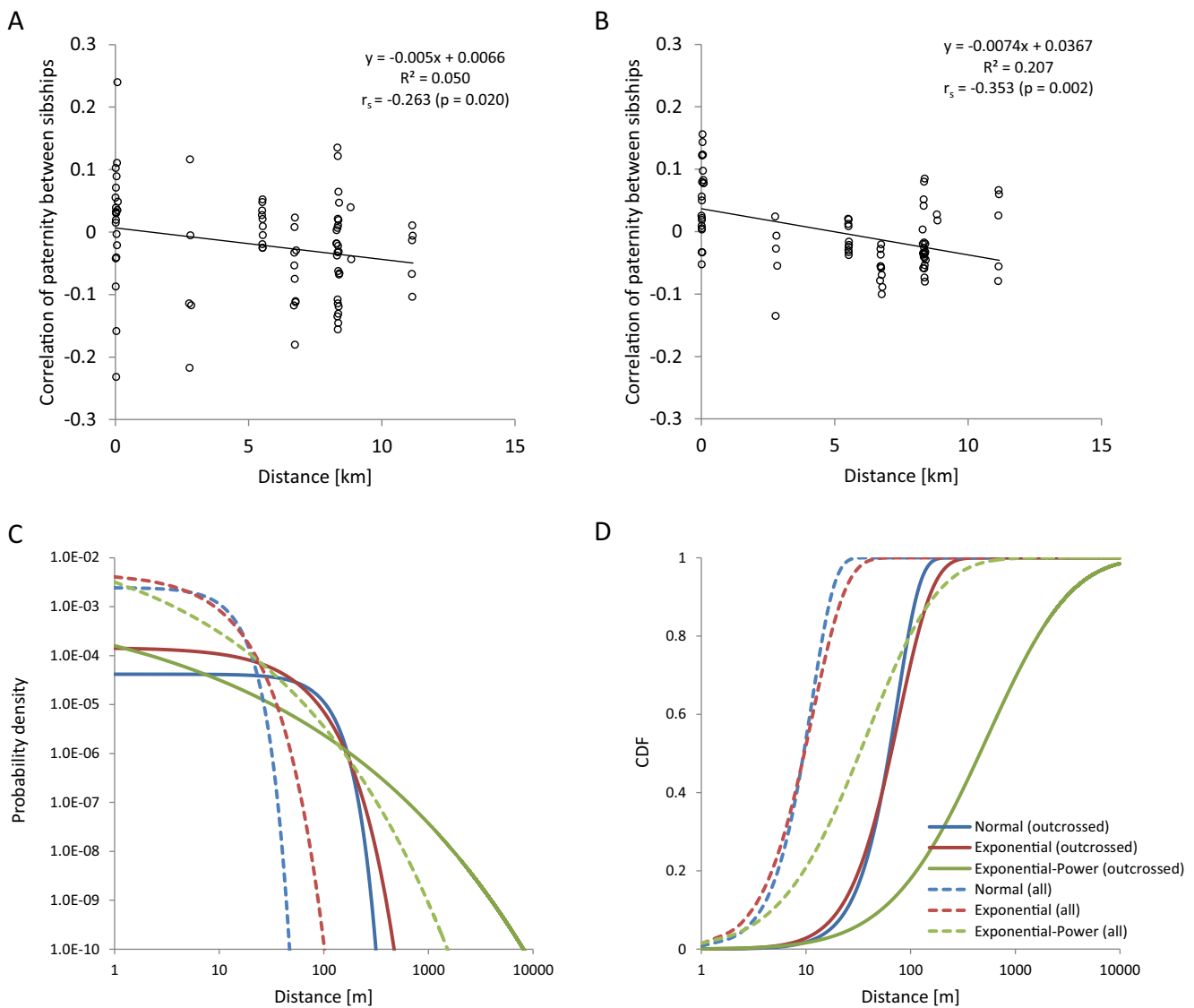
One advantage of the algorithm described above was the possibility to estimate the posterior probability that the  $j$ -th individual in the  $i$ -th family was produced through outcrossing ( $P_{\text{out}}$ ).  $P_{\text{out}}$  was estimated as the mean  $X_{ij}$  over MCMC cycles, disregarding the first 100,000 values for burn-in. Finally, every progeny individual with  $P_{\text{out}} > 0.95$  was classified as being outcrossed. The Bayesian approach was implemented in the computer program written in Delphi/Pascal (CEMBRA program is available at [http://www.ukw.edu.pl/pracownicy/strona/igor\\_chybicki/software\\_ukw/](http://www.ukw.edu.pl/pracownicy/strona/igor_chybicki/software_ukw/)).

### Analysis of pollen dispersal

Because only the maternal families were available, we used the indirect approach called KinDist (Robledo-Arnuncio et al. 2006) to make inferences about pollen dispersal kernel. KinDist uses the theoretical relationship between the pollen dispersal kernel and the standardized kinship between pollen gametes, inferred from an offspring's genotype after subtracting the maternal input. This approach was chosen because, unlike TwoGener (Smouse et al. 2001), it does not require an effective density of pollen parents ( $D_e$ ) to be known. We need to stress here that in our case, due to problematic delimitation of a mating population (see Dzialuk et al.

2014),  $D_e$  cannot be practically assessed. During the analysis, we focused particularly on the assessment of whether the dispersal kernel is thin- or fat-tailed or whether long-distance pollen dispersal (LDD) is a likely phenomenon in the study system. Therefore, we assumed that pollen dispersal follows an exponential family kernel (the exponential-power kernel; Austerlitz et al. 2004), defined in terms of the parameters: the shape ( $b$ ) and the scale ( $a$ ). The exponential-power kernel is known for sufficient flexibility in terms of kurtosis (Austerlitz et al. 2004). Generally, if  $b > 1$  then the kernel is considered as thin-tailed ( $b = 2$  leads to the normal distribution). This kind of kernels is characterized by a very low probability of LDD. On the other hand, if  $b < 1$  then the kernel is considered as fat-tailed. In this case, the probability of LDD is relatively high. Therefore, in order to infer the potential for LDD, we compared the three models: normal, exponential, and exponential-power (fixing  $b$  at 2, 1 and treating it as an estimable parameter, respectively). The analysis was performed using the POLDISP package (Robledo-Arnuncio et al. 2007). According to the recommendation of the authors (Robledo-Arnuncio et al. 2007), at the initial step, we checked whether pair-wise paternity correlation coefficients decreased significantly with distance. For this purpose, the Spearman rank correlation coefficient was used with  $p$  values estimated using the permutation procedure. Also, the threshold distance between mother trees, at which there is likely no correlation in paternity between sibships (Robledo-Arnuncio et al. 2006), was set to the value of 6 km. This value was equal to the distance at which the linear regression function of the observed correlation of paternity started to be negative (Fig. 2b). Also, the distance of 6 km corresponds roughly to the maximum distance between nearest populations in this study. Nonetheless, setting the threshold distance to 1 km, i.e., treating between-populations pollen pools uncorrelated a priori, did not change the results (not shown).

Because POLDISP offers no explicit tool for model comparison, we used the general theory developed for a non-linear regression (Burnham and Anderson 2002). First, for the  $i$ th model we estimated the log-likelihood, as  $\ell_i(\hat{\theta}) = -\frac{1}{2}n(\log(\hat{\sigma}_i^2) + \log(2\pi) + 1)$ , where  $\hat{\sigma}_i^2$  was estimated as the sum of squares of residuals (SSR) divided by the number of



**Fig. 2** Genetic structure of pollen gametes (pooled data) and the inferred pollen dispersal based on the Kindist approach. The *upper panel* shows correlation of paternity as a function of distance between mother trees for the entire sample of progeny (**a**) and the subsample of outcrossed progeny (**b**). The best fitting regression lines together with the Spearman rank correlation coefficients are also shown. The *lower panel* shows pollen dispersal kernel (probability density; given on a log scale; **c**) and the

cumulative probability distribution function (CDF; **d**) of dispersal distance (given on a log scale) estimated assuming the three models of dispersal. *Dashed lines* show the results based on the entire sample, while *solid lines* show the results based on the outcrossed progeny. Note that according to the model comparison procedure, the exponential-power kernel showed the best fit (see the text for details)

data points for the *i*th dispersal model (SSR obtained as the standard output of POLDISP). Subsequently, we estimated Akaike Information Criterion values as  $AIC_i = -2\ell_i(\hat{\theta}) + 2K$ , where *K* equalled the number of estimated parameters for the *i*th model (including  $\hat{\sigma}_i^2$ ). The model with the smallest  $AIC_i$  was considered the best one.

Both TwoGener and KinDist work under the assumption of random (or null) selfing (Austerlitz et al. 2004; Robledo-Arnuncio et al. 2006; Carpentier 2011). Because in our case this assumption was generally violated (see the “Results” section), the above estimation procedures were applied to

two data sets. First, all progeny individuals were used as data (the untrimmed sample). In the second case, only individuals produced through apparent outcrossing were used. They were identified using the Bayesian approach described above.

## Results

### Genetic polymorphism

In total, 97 alleles were detected in progeny genotypes, ranging from 4 to 23 per locus (data not shown). Overall

expected heterozygosities varied from 0.164 to 0.849, with the average 0.562. Observed heterozygosities spanned between 0.144 and 0.705, with the average 0.394. We observed significant deficiency of heterozygotes for all markers. The multilocus exclusion probability equaled 0.98499.

#### Average outcrossing rates

The overall multilocus outcrossing rate  $t_m$  estimated jointly for all mother trees and seasons equaled 0.725 and was significantly different from both extremes (0 and 1; Table 3). When estimated separately for small and large populations, i.e., ignoring the seasonal effect, the average  $t_m$  equaled 0.513 and 0.886, respectively. For comparison,  $t_m$  values averaged per sampling year showed little variation, ranging from 0.661 to 0.773. However, the season-specific outcrossing rates did not account for the effect of population size. The overall single-locus outcrossing rate equaled 0.593 and was significantly different from both 0 and 1. For small and large populations, the average  $t_s$  equaled 0.404 and 0.769, respectively, while it spanned between 0.519 and 0.679 among sampling years. All single-locus outcrossing rates were significantly lower as

compared with the parallel multilocus indices (Table 4). The overall correlation of selfing among loci equaled 0.673 ( $\pm 0.056$ ). Large subpopulations had significantly higher  $r_s$  than small subpopulations. At the seasonal level,  $r_s$  ranged between 0.644 (2010) and 0.691 (2009). All  $r_s$  estimates deviated significantly from both 0 and 1.

In addition, we performed comparisons between population size-specific or year-specific estimates and the overall average (obtained for the entire data set) using the Z test (and bootstrap standard errors). The procedure revealed that  $t_m$  and  $t_s$  for large and small populations were significantly higher and lower, respectively, as compared with the overall averages. Also,  $r_s$  for large subpopulations was significantly lower than the overall average.

Mating system parameters were also estimated grouping data simultaneously per year and population size. These estimates revealed that seeds sampled in 2009 and 2010 showed consistently the lowest and the highest outcrossing rates, respectively, regardless of population size. This suggests that mating system parameters might reveal seasonality. Nonetheless, based on the simple statistical comparisons, we were unable to confirm the significance of this pattern (Table 3).

**Table 3** Estimates of outcrossing rate together with standard errors (SE)

Estimation level	Parameter							
	$t_m$	SE	$t_s$	SE	$t_m - t_s$	SE	$r_s$	SE
Overall	0.725	0.069	0.593	0.071	0.132	0.015	0.673	0.056
Per-size <sup>a</sup>								
Small	0.513*	0.085	0.404*	0.074	0.108	0.023	0.703	0.082
Large	0.886*	0.024	0.769*	0.034	0.118	0.019	0.485*	0.069
Per-year <sup>b</sup>								
2008	0.773	0.072	0.628	0.077	0.145	0.019	0.645	0.084
2009	0.661	0.101	0.519	0.100	0.142	0.024	0.691	0.089
2010	0.752	0.106	0.679	0.112	0.074*	0.023	0.644	0.099
Per-size/per-year								
Small/2008	0.580	0.139	0.447	0.137	0.133	0.028	0.684	0.123
Small/2009	0.352	0.180	0.251	0.134	0.101	0.053	0.723	0.179
Small/2010	0.642	0.145	0.574	0.143	0.068	0.033	0.596	0.184
Large/2008	0.901*	0.042	0.770*	0.042	0.131	0.019	0.481	0.158
Large/2009	0.863*	0.037	0.740*	0.050	0.123	0.031	0.420*	0.093
Large/2010	0.909	0.151	0.853	0.173	0.056	0.050	0.814	0.231

Asterisks denote significant difference ( $\alpha=0.05$ ; after the Šidák correction for multiple tests) between per-size or per-year averages and the overall average

$t_m$  the multi-locus outcrossing rate,  $t_s$  the single-locus outcrossing rate,  $r_s$  the correlation of selfing among loci

<sup>a</sup> Grouping by population size disregarded seasonal variation of outcrossing rates (Small populations: Dolina Białego, Dolina Sucheje Kasprowej; Large populations: Dolina Waksmundzka, Morskie Oko)

<sup>b</sup> Grouping by collection year disregarded variation in outcrossing rates due to population size



Effects of population size and seasonal variation and individual outcrossing probabilities

Six candidate regression models were used to explain the observed variation in outcrossing rates (Table 4). The *DBH* model (DIC=20644) showed the quality comparable to that of the null model (DIC=20646). Although the *Season* model (DIC=20640) showed the significant improvement of fit as compared to the null model, it was apparently worse, as compared to the *Population size* model (DIC=20458). This result was in agreement with the result of simple comparisons of mean outcrossing rates, described above. The model accounting for both seasonality and the effect of population size revealed the best quality (DIC=20439), indicating that both factors had confounding effect on the observed outcrossing rates. The addition of *DBH* did not improve fit substantially ( $\bar{D}$ ), decreasing at the same time quality (DIC) of the model. Moreover, the credible interval for the slope  $\gamma$  included 0, indicating that outcrossing rates were not associated with *DBH* of mother trees.

Based on the best model (*Population size+season*), in the case of the *Season* variable, the slope coefficient  $\beta_{s2}$  for the season 2009 was significantly lower than 0. It thus indicated that outcrossing rate, or more precisely:

outcrossing/selfing odds, was significantly lower in the subsample of seeds collected in that season compared to the average. Taking  $\exp(\beta_{s2})$ , one may predict that the outcrossing/selfing odds relative to the geometric mean odds was about 0.56 in 2009. Similarly,  $\beta_{p1}$  for the small population size was significantly negative, indicating that outcrossing rates were significantly lower in seeds collected in small populations than in seeds taken from large populations. In this case, the outcrossing/selfing odds ratio for small populations equalled 0.34.

As a by-product of the Bayesian estimation, a series of posterior probabilities for progeny of being produced through outcrossing ( $P_{out}$ ) was obtained. Using a criterion of  $P_{out} > 0.95$ , we separated 812 (out of 1129) individuals that were produced through apparent outcrossing (Table 1). Among 317 individuals classified as selfed, the posterior probability of being produced through outcrossing ranged from 0 to 0.794, with the average 0.029. In other words, the Bayesian procedure only exceptionally resulted in ambiguous paternities and the vast majority of individuals were classified as being selfed or outcrossed with nearly 100 % certainty. It is worth mentioning that the model with and without over-dispersion in family outcrossing rates gave exactly the same number of selfed vs. outcrossed individuals.

**Table 4** Effects of population size and seasonal variation on outcrossing rates

Model	Estimate	Regression parameters							$\bar{D}$	$p_D$	DIC
		$\alpha$ Intercept	$\beta_{s1}$ 2008	$\beta_{s2}$ 2009	$\beta_{s3}$ 2010	$\beta_{p1}$ Small	$\beta_{p2}$ Large	$\gamma$ DBH			
Population size+season+ DBH (the full model)	Mean	0.642	0.132	-0.562	0.430	-1.071	1.071	0.011	20,356	84.7	20,441
	CI <sub>2.5%</sub>	-0.393	-0.179	-0.865	-0.037	-1.338	0.819	-0.009			
	CI <sub>97.5%</sub>	1.674	0.425	-0.267	0.949	-0.817	1.338	0.032			
Population size+season	Mean	1.172	0.145	-0.572	0.428	-1.074	1.074		20,357	82.4	20,439
	CI <sub>2.5%</sub>	0.963	-0.147	-0.868	-0.036	-1.323	0.829				
	CI <sub>97.5%</sub>	1.399	0.440	-0.284	0.916	-0.828	1.324				
Population size	Mean	1.054				-1.005	1.005		20,381	77.8	20,458
	CI <sub>2.5%</sub>	0.894				-1.252	0.770				
	CI <sub>97.5%</sub>	1.236				-0.770	1.252				
Season	Mean	1.009	0.220	-0.339	0.119				20,560	79.7	20,640
	CI <sub>2.5%</sub>	0.836	-0.046	-0.592	-0.316						
	CI <sub>97.5%</sub>	1.193	0.488	-0.087	0.556						
DBH	Mean	-0.021						0.021	20,566	78.4	20,644
	CI <sub>2.5%</sub>	-0.767						0.006			
	CI <sub>97.5%</sub>	0.723						0.036			
Null	Mean	0.968							20,570.9	75.5	20,646
	CI <sub>2.5%</sub>	0.839									
	CI <sub>97.5%</sub>	1.097									

Model, the regression model for outcrossing rates; Estimate, the Bayesian estimate (mean – the posterior mean; CI<sub>2.5%</sub> and CI<sub>97.5%</sub>, bounds of the 95 % credible interval); Regression parameters, the intercept ( $\alpha$ ) and the slopes of the logistic regression function ( $\beta_{sm}$ , the regression slope coefficient for the effect of the *m*-th level in the *season* variable;  $\beta_{pn}$ , the regression slope coefficient for the effect of the *n*-th level in *population size* variable;  $\gamma$ , the regression slope coefficient for the effect of *DBH*);  $\bar{D}$  the measure of the model fit;  $p_D$ , the effective number of parameters in the model; DIC, the Deviance Information Criterion measuring the model quality

## Pollen dispersal kernel

Generally, we found the significant decrease in correlation of paternity between sib-ships with a distance (Fig. 2). However, when outcrossed seeds were only retained, the association between the correlation of paternity and the distance appeared to be stronger (the Spearman rank correlation  $r_s = -0.353$ ;  $p$  value = 0.002) than for the total sample of progeny ( $r_s = -0.263$ ;  $p$  value = 0.020).

When all pollen gametes (i.e., all progeny) were used in the analysis, the estimated average pollen dispersal distance varied from 10.1 to 69.1 m, depending on the shape of the exponential-power function (Table 5). According to the model comparison procedure, the full exponential-power kernel (i.e., for which both scale and shape parameters were set estimable) appeared the optimal one, with the Akaike weight equal to 59.3 %. For comparison, the normal kernel had the lowest Akaike weight of 12.8 %. The median of dispersal distance reached up to 32 m (Fig. 2), suggesting that pollen dispersal is restricted mostly to a local neighborhood of a mother tree.

Generally, the analysis based on the outcross pollen gametes gave the scale of pollen dispersal one order of magnitude larger compared with the untrimmed procedure. The mean and median dispersal distance ranged between 66.1 and 1,266.5 and between 62.2 and 458.8 m, respectively (Table 5). However, also in this case, the full exponential-power kernel appeared the optimal one, with the Akaike weight of 99.4 %. Interestingly, the shape parameter  $b = 0.206$  was similar to that obtained based on the untrimmed procedure.

## Discussion

We showed that *P. cembra* in the Tatra Mts. is characterized by the mixed mating system. The average outcrossing rate of 72 % is comparable to results for populations from the Eastern Carpathians (Politov and Krutovskii 1994; Politov et al. 2008), while slightly lower than outcrossing rates found for populations in the Alps (81 %; Lewandowski and Burczyk 2000; Salzer and Gugerli 2012). Furthermore, using the newly developed method, we detected confounding effects of population size and seasonal variation on outcrossing rates. In concordance with the expectations, small populations were characterized by lower outcrossing rate. However, we found no association between DBH of a mother and the outcrossing rate. Also, our study demonstrated that biparental inbreeding is a significant component of mating system. Interestingly, biparental inbreeding was found in both small and large populations. Finally, we inferred that pollen dispersal follows a heavy-tailed distribution, enabling long-distance pollen dispersal.

Density-dependent mating system is well documented in the literature on plant mating systems. Most often, low stand density significantly increases the proportion of selfed progeny (Rajora et al. 2002; Robledo-Arnuncio et al. 2004; Mimura and Aitken 2007). Comparing core and peripheral populations in the Swiss Alps, Salzer and Gugerli (2012) found sharp differences in selfing rates (9 % vs. 35 %), indicating that pollen availability must be negatively associated with population size in *P. cembra*. Other reports for *P. cembra* seem to be in agreement with this statement (Lewandowski and Burczyk 2000; Politov et al. 2008). In our case, significant differences in outcrossing rates found between small and large

**Table 5** Estimates of the pollen dispersal kernel

Dispersal kernel	<i>a</i>	<i>b</i>	$\delta$	Q <sub>50%</sub>	RSS	$\ell_i(\hat{\theta})$	<i>AIC<sub>i</sub></i>	<i>w<sub>i</sub></i>
Total progeny								
Normal	11.40	2 <sup>a</sup>	10.1	9.5	5.670	-8.44	20.88	0.128
Exponential	5.73	1	11.5	9.6	5.559	-7.67	19.33	0.279
Exponential power	0.0259	0.273	69.1	32.4	5.314	-5.91	17.82	0.593
Outcrossed progeny only								
Normal	87.39	2	66.1	62.2	10.174	-31.24	66.48	0.002
Exponential	33.07	1	77.4	65.0	9.979	-30.48	64.97	0.004
Exponential power	0.0085	0.206	1,266.5	458.8	8.447	-23.98	53.97	0.994

*a*, the scale parameter of the dispersal kernel; *b*, the shape parameter of the dispersal kernel;  $\delta$ , the mean distance of pollen dispersal (in meters); Q<sub>50%</sub>, the median distance of pollen dispersal (in meters); RSS, the sum of squares of residuals;  $\ell_i(\hat{\theta})$ , the likelihood of the parameter estimates; *AIC<sub>i</sub>*, Akaike Information Criterion; *w<sub>i</sub>*, the Akaike weight for the *i*-th model

<sup>a</sup> Italicized values denote that these parameters were treated as constants during estimation

populations in the Tatras are in line with these conclusions. According to the reproductive assurance hypothesis (Lloyd 1979), when outcross pollen is unavailable, selfing allows to continue seed production. In our case, small subpopulations of *P. cembra* in the Tatras can experience outcross pollen limitation because sporadic mature trees are scattered in the forest so that inter-individual distances (even between nearest neighbors) are longer, and due to the presence of the other species, harder to traverse (Milleron et al. 2012) as compared with large, dense subpopulations (especially Morskie Oko).

The average outcrossing rate found in this study translates into the expected inbreeding level  $F=(1-0.72)/(1+0.72)=0.16$ . In the Ukrainian Carpathians, Politov et al. (2008) found a good correspondence between outcrossing rates and inbreeding levels estimated at embryonic stage. However, low outcrossing rates do not necessarily lead to high inbreeding at the adult stage. In *Abies balsamea*, the population at the highest elevation revealed the lowest outcrossing rate measured at the seed stage (Neal and Adams 1985). However, the adult population showed an excess of heterozygotes, suggesting that selection favors heterozygotes in this marginal environment. The results for *P. cembra* fully support this prediction (Salzer and Gugerli 2012) in respect to populations in the Swiss Alps. In the case of the Tatras, a related study revealed that the population of mature individuals is only slightly inbred, with the overall average  $F=0.051$  (Dzialuk et al. 2014). Using the equilibrium equation (Ritland 1990), we can estimate the relative fitness of selfed individuals to be as low as  $w = 2tF/((1-t)(1-F))=0.283$ . Hence, it seems reasonable to expect that inbred individuals will be outcompeted by non-inbred ones during the seedling development, as often observed in pine species (Savolainen and Hedrick 1995; Koelewijn et al. 1999; Bower and Aitken 2007). On the other hand, the evolutionary theory predicts that recurrent selfing accompanied by intensive selection against inbred individuals can lead to genetic purging (Crnokrak and Barrett 2002). In such conditions, a number of deleterious alleles is significantly reduced, opening possibility for increased self-fertilization and inbreeding without negative consequences for fitness. An experimental study on Scots pine provided good support for this theory (Kärkkäinen et al. 1996). In our case, small and isolated populations of *P. cembra* seem to be particularly prone to genetic purging, assuming they will continue to experience outcross pollen limitation.

It should be stressed that the reported selfing rates can be underestimated, because our estimates did not account for seed abortion, which is highly correlated with self-fertilization (Koski 1971; Savolainen et al. 1992; Kärkkäinen et al. 1999). The total probability of survival of a seed can be expressed as  $P(B)=sw_s+(1-s)w_t$ , where  $s$  is the actual self-fertilization rate,  $w_s$  and  $w_t$  is the proportion of progeny survived after selfing and outcrossing, respectively. After Ritland (1990), let  $w=w_s/w_t$ . Then, the (posterior)

probability for a seed to be after self-fertilization, given that it survived, is equal  $P(A|B) = \frac{sw}{1-s(1-w)}$ . In fact,  $P(A|B)$  is equivalent to the effective self-fertilization rate  $\hat{s}$  typically estimated based on genotypes of viable progeny. Consequently, the true selfing rate in small and large populations in the Tatras adjusted for seed abortion, i.e., estimated as  $s = \frac{\hat{s}}{s+w(1-\hat{s})}$ , can be as high as 0.788 and 0.313, respectively.

Interestingly, these estimates are in a good agreement with equivalent estimates for peripheral and core populations in the Alps (Salzer and Gugerli 2012; see the Appendix). According to the mass action model (Holsinger 1991), where the frequency of self-fertilization is the proportion of self pollen in the total pollen received by a given tree, in small populations only 21 % of the pollen pool of a given tree is not produced by that tree. For comparison, in large populations the outcross pollen is expected to prevail (69 % of the total). It thus seems that pollen limitation is among the main threats in small subpopulations of *P. cembra*.

Knowledge about pollen dispersal in *P. cembra* is extremely scarce. Using a parentage assignment approach, Salzer (2011) revealed that within a marginal stand of *P. cembra* in the Alps, effective pollen dispersal distance was, depending on the conditions, <200 m or about 315 m on average (maximum distance of 659 m). Our own analysis revealed that the total pollen dispersal follows a heavy-tailed dispersal kernel, confirming great potential of *P. cembra* to long-distance pollen dispersal. The mean dispersal of about 1 km inferred in this study was threefold higher than previously found. However, estimates reported by Salzer (2011) did not account for pollen immigration, which reached about 40 % from at least 1 km. Thus, the total effective pollen dispersal in the Alps could be higher. Using the cumulative probability distribution (CDF) for the exponential-power kernel (Fig. 2) estimated in this study, we can infer that in the Tatras about 69 % pollen travels up to 1 km, while 31 % would be immigrating from at least 1 km. Thus, our results are in line with the findings based on the direct approach by Salzer (2011). They also correspond well with general knowledge that species with saccate pollen are characterized by great pollen dispersal potential (Schwendemann et al. 2007). In the case of *P. cembra*, the potential can be further increased by suitable environmental conditions, such as high wind speed. Nonetheless, at a regional scale, the specific configuration of *P. cembra* populations would eventually determine the potential for long-distance dispersal. More detailed, parentage-based study would be helpful to compare small and large populations in this respect.

An important component of plant mating system is mating between relatives, leading to biparental inbreeding (e.g., Degen et al. 2004; Hirao 2010; Fenster et al. 2003). Interestingly, except for our study, no strong signatures of biparental inbreeding in *P. cembra* have been shown (Lewandowski and Burczyk 2000; Politov et al. 2008). The

between-locus correlation of selfing ( $r_s$ ) can be interpreted as roughly equal to the proportion of inbreeding due to selfing, while  $(1-r_s)$  as a proportion of inbreeding due to mating between relatives (Ritland 2002). Consequently, small subpopulations in the Tatras showed a lower proportion of mating between relatives as compared with large ones. However, a lower proportion of consanguineous mating is somewhat counterintuitive, because typically one can expect that biparental inbreeding tends to increase as a number of mates decreases (Morgante et al. 1991; Takayama and Isoagi 2005). One possible scenario can be proposed taking into account the result that pollen dispersal was large. In this case, because short distances are under-represented in small subpopulations (due to lower density), it seems very likely that most outcross pollen travels long distances, favoring gene flow between unrelated individuals. Similar observations were previously made for isolated trees of *Picea glauca* (O’Connell et al. 2007). On the contrary, in large and denser subpopulations, where short and intermediate distances are well represented, mating between close neighbors can occur relatively frequently (see Fig. 2). However, in such conditions, biparental inbreeding can evolve only if relatives tend to be spatially closer than at random (Fenster et al. 2003). In the case of *P. cembra*, seeds are typically dispersed by birds (*N. caryocatactes*; Tomback et al. 1993) and squirrels (*Sciurus vulgaris*). Given that seeds of *P. cembra* have no structures facilitating dispersal, such a mechanism allows efficient colonization of new habitats. Birds typically bury several seeds together (Tomback et al. 1993), enabling the development of a family structure. In fact, strong spatial genetic structure was found in populations of *P. cembra* in the Swiss Alps (Salzer 2011), providing some support for this prediction. However, we cannot be sure that strong spatial genetic structure is also present in the study populations, where clumps of individuals due to seed caching seem to be relatively rare.

Temporal variation in outcrossing rates is often found in natural conifer populations (e.g., Shea 1987; Cheliak et al. 1985; El-Kassaby et al. 1993). It is due to various reasons, with environmental variation being most often indicated. However, seasonality in outcrossing rates inferred in our study might be also caused by the masting behaviour of the species (Ulber et al. 2004). In fact, the significantly lower outcrossing rate in 2009 coincided with the large number of seeds collected in that year, relative to the other years (see M&M). Although we are rather skeptical about a definitely causal linkage between these two characteristics, this might be reasonable to expect that higher fecundity leads to lower outcrossing rate (Shea 1987). Nonetheless, similarly as in O’Connell et al. (2006), we failed to show that a tree size, used as a proxy for fecundity, is associated with outcrossing rates. Therefore, further studies are needed to confirm the role of both male fecundity and masting behavior in shaping mating system parameters of *P. cembra*.

Usually, *P. cembra* in the Tatras occupies hardly accessible locations. In consequence, our study was based on a limited number of mother trees and localities. Despite that inferences on mating system parameters were in concordance with predictions, we need to stress that a small number of sampled mother trees may preclude very strong conclusions, especially regarding conservation strategies for the species. A special caution should be paid in the case of pollen dispersal inferences. In order to achieve sufficient statistical power, the parameters of pollen dispersal kernel were estimated based on the entire sample, without any respect of temporal or spatial variation of the dispersal, which might be present in our case. Thus, the estimated dispersal kernel represents rather the average across different population types and different seasons. On the other hand, the indirect approach (KINDIST) used in this study requires correlation of paternity to be a continuous and significant function of a distance between mother trees (Robledo-Arnuncio et al. 2006, 2007). Moreover, the latter is assumed to result entirely from pollen dispersal and not population genetic structure (Austerlitz and Smouse 2001). The absence of correlation between genetic and geographic distance found in the Tatras (Dzialuk et al. 2014) suggests that the observed pollen structure was indeed a result of limited pollen dispersal. Nonetheless, due to discontinuous sampling (see Fig. 2), the estimation procedure might suffer from the bias, which cannot be assessed based on the existing data.

As a threatened species, *P. cembra* in the Tatras is under strict protection. However, in situ conservation has been questioned as an efficient and satisfactory means to preserve gene pools. Our results show that gene flow through pollen should be generally sufficient to counterbalance genetic drift. In fact, studying genetic structure of *P. cembra* in the Tatras, we detected very weak genetic differentiation (Dzialuk et al. 2014), suggesting that gene exchange between subpopulations is probably high. On the other hand, based on mating system parameters, we argued that small and isolated subpopulations of *P. cembra* both in the Alps and Carpathians can experience severe local pollen limitation, which can dramatically lower seed production (Salzer and Gugerli 2012) and fitness of individuals due to inbreeding. In particular, seed yield can be reduced (Salzer and Gugerli 2012), lowering (variance) effective population size. However, because tree species seem to be generally “resistant” to the impact of fragmentation on the genetic diversity (O’Connell et al.

**Table 6** Seed abortion and self-fertilization rate estimated for *P. cembra* in the Alps (based on Salzer and Gugerli 2012)

Feature	Parameter	Peripheral	Core
Seed abortion rate	$1-P(B)$	0.4371	0.0853
Self-fertilization rate	$\hat{s}$	0.3515	0.09325

2006; Kramer et al. 2008), the question of whether high adult genetic homogeneity in the Tatras (Dzialuk et al. 2014) will be preserved in future generations remains open. To answer this, we need to know if potentially high pollen-mediated gene flow can preclude erosion of the gene pool in small populations. Therefore, the next-generation genetic structure remains a pending issue for future studies.

**Acknowledgments** This study was supported by the research grant from the Polish National Science Centre (NN304 129336) to AD. We are grateful to the Tatra National Park, and especially to Mr. Tomasz Mączka, for his help in the field work. We thank Katarzyna Meyza and Ewa Sztupecka for their assistance in the laboratory work, Stewart Berlocher for his help with English editing, and three reviewers for their constructive comments on the first version of the manuscript.

**Conflict of interest** The authors declare no conflict of interest.

**Data archiving** If the manuscript is accepted, genotype data will be submitted to Dryad. All genotype data are deposited at the Reaserch Gate ([www.researchgate.net](http://www.researchgate.net), doi:10.13140/2.1.1198.2727).

## Appendix

In the discussion we showed that knowing the relative fitness of embryos after selfing ( $w$ ) and the effective self-fertilization rate ( $\hat{s}$ ), we can estimate the true self-fertilization rate as

$= \frac{\hat{s}}{s - \alpha(1 - \hat{s})}$ . Here we use our considerations to interpret the

results on *Pinus cembra* reported in Salzer and Gugerli (2012). The authors showed that both the rate of seed abortion ( $P(B)$ ) and the selfing rate ( $\hat{s}$ ) in peripheral populations was higher than in core ones. Using subscripts  $p$  and  $c$  for peripheral and core populations, respectively, and substituting

$\frac{\hat{s}}{s - w(1 - \hat{s})}$  for  $s$ , we can express the ratio of survival rates in

both population types as  $\frac{P(B_p)}{P(B_c)} = \frac{w_c + \hat{s}_c(1 - w_c)}{w_p + \hat{s}_p(1 - w_p)}$ . Now, assuming

that there is no differences between survival rates in both population types, i.e.,  $w_c = w_p = w$ , we can rearrange the last

equation in order to compute  $w$ , i.e.,  $w = \frac{P(B_c)\hat{s}_c - P(B_p)\hat{s}_p}{P(B_p)(1 - \hat{s}_p)} - P$

$(B_c)(1 - \hat{s}_c)$ . Using the data from Salzer and Gugerli (2012) (Table 6), we obtain  $w$  to be equal 0.242. Also, knowing  $w$  we

can compute the actual selfing rate (i.e., the parameter value which accounts for seed abortion) in core and peripheral populations in the Alps to be 0.300 and 0.691.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. III. Impact of adult population structure. *Genet Res Camb* 78:271–280
- Austerlitz F, Dick CW, Dutech C, Klein EK, Oddou-Muratorio S, Smouse PE, Sork VL (2004) Using genetic markers to estimate the pollen dispersal curve. *Mol Ecol* 13:937–954
- Bower AD, Aitken SN (2007) Mating system and inbreeding depression in whitebark pine (*Pinus albicaulis* Engelm.). *Tree Genet Genome* 3:379–388
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York, p 488
- Carpentier F (2011) Modélisations de la dispersion du pollen et estimation à partir de marqueurs génétiques. The PhD Thesis, Université Montpellier II
- Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. *Annu Rev Ecol Syst* 18:237–268
- Cheliak WM, Dancik BP, Morgan K, Yeh FC, Strobeck C (1985) Temporal variation of the mating system in a natural population of jack pine. *Genetics* 109:569–584
- Chmiel J (1996) Naturalny zasięg a rozsiedlenie limby *Pinus cembra* L w Tatrach Polskich. In: A Kownacki (Ed) Przyroda Tatrzaskiego Parku Narodowego a człowiek—2 Biologia, pp 63–66 (in Polish)
- Chybicki IJ (2013) Note on the applicability of the F-model in analysis of pollen pool heterogeneity. *J Hered* 104:578–585
- Chybicki IJ, Burczyk J (2013) Seeing the forest through the trees: comprehensive inference on individual mating patterns in the mixed stand of *Quercus robur* and *Q. petraea*. *Ann Bot* 112:561–574
- Coelho ASG, Vencovsky R (2003) Intrapopulation fixation index dynamics in finite populations with variable outcrossing rates. *Sci Agric* 60:305–313
- Crnokrak P, Barrett SCH (2002) Perspective: purging the genetic load: a review of the experimental evidence. *Evolution* 12: 2347–2358
- Degen B, Bandou E, Caron H (2004) Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana. *Heredity* 93:585–91
- Doyle J, Doyle J (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Dzialuk A, Chybicki I, Gout R, Mączka T, Fleischer P, Konrad H, Curtu AL, Sofletea N, Valadon A (2014) No reduction in genetic diversity of Swiss stone pine (*Pinus cembra* L.) in Tatra Mountains despite high fragmentation and small population size. *Conserv Genet*. doi: 10.1007/s10592-014-0628-6
- El-Kassaby YA, Jaquish B (1996) Population density and mating patterns in western larch. *J Hered* 87:438–443
- El-Kassaby Y, Meagher MD, Davidson R (1993) Temporal variation in the outcrossing rate in a natural stand of western white pine. *Silvae Genet* 42:131–135
- Farris MA, Mitton JB (1984) Population density, outcrossing rate, and heterozygote superiority in ponderosa pine. *Evolution* 38:1151–1154
- Fenster CB, Vekemans X, Hardy OJ (2003) Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (Leguminosae). *Evolution* 57:995–1007
- Goodwillie C, Kalisz S, Eckert C (2005) The evolutionary enigma of mixed mating in plants: occurrence, theory and empirical observations. *Annu Rev Ecol Evol Syst* 36:47–79
- Hirao AS (2010) Kinship between parents reduces offspring fitness in a natural population of *Rhododendron brachycarpum*. *Ann Bot* 105: 637–646
- Hoff PD (2009) A first course in Bayesian statistical methods. Springer, New York

- Holsinger KE (1991) Mass-action models of plant mating systems: the evolutionary stability of mixed mating systems. *Am Nat* 138:606–622
- Holsinger KE (1996) Pollination biology and the evolution of mating systems in flowering plants. *Evol Biol* 29:107–149
- Hosmer DW, Lemeshow S (2000) Applied logistic regression, 2nd edn. Wiley, New York
- Jain SK (1976) The evolution of inbreeding in plants. *Annu Rev Ecol Syst* 7:469–495
- Jain SK, Marshall RD (1985) Simulation of models involving mixed selfing and random mating. I. Stochastic variation in outcrossing and selection parameters. *Heredity* 23:411–432
- Jamieson A, Taylor SC (1997) Comparisons of three probability formulae for parentage exclusion. *Anim Genet* 28:397–400
- Jamnicky J (1981) Rosirenica staw borowice limby (*Pinus cembra* L.) o Zapadnych Karpatoch. *Zb. Praca TANAPu* 22:5–20 (in Slovakian)
- Kamm U, Gugerli F, Rotach P, Edwards P, Holderegger R (2011) Genetic consequences of low local tree densities—implications for the management of naturally rare, insect pollinated species in temperate forests. *Forest Ecol Manag* 262:1047–1053
- Kärkkäinen K, Koski V, Savolainen O (1996) Geographical variation in the inbreeding depression of Scots pine. *Evolution* 50:111–119
- Kärkkäinen K, Savolainen O, Koski V (1999) Why do plants abort so many developing seeds: bad offspring or bad maternal genotypes? *Evol Ecol* 13:305–317
- Knowles P, Fumier GR, Alesiuk MA, Perry DJ (1987) Significant levels of self-fertilization in natural populations of tamarack. *Can J Bot* 65:1087–1091
- Koelwijn HP, Koski V, Savolainen O (1999) Magnitude and timing of inbreeding depression in Scots pine (*Pinus sylvestris* L.). *Evolution* 53:758–768
- Koski V (1971) Embryonic lethals of *Picea abies* and *Pinus sylvestris*. *Community Inst For Fenn* 75:1–30
- Kramer AT, Ison JL, Ashley MV, Howe HF (2008) The paradox of forest fragmentation genetics. *Conserv Biol* 22:878–885
- Lande R, Schemske DW (1985) The evolution of self-fertilization and inbreeding depression in plants. I. Genetics models. *Evolution* 39:24–40
- Lewandowski A, Burczyk J (2000) Mating system and genetic diversity in natural populations of European larch (*Larix decidua*) and stone pine (*Pinus cembra*) located at higher elevations. *Silvae Genet* 49:158–161
- Lloyd DG (1979) Some reproductive factors affecting the selection of self-fertilization in plants. *Am Nat* 113:67–79
- Milleron M, Lopez de Heredia U, Lorenzo Z, Perea R, Dounavi A, Alonso J, Gil L, Nanos N (2012) Effect of canopy closure on pollen dispersal in a wind-pollinated species (*Fagus sylvatica* L.). *Plant Ecol* 213:1715–1728
- Mimura M, Aitken SN (2007) Increased selfing and decreased effective pollen donor number in peripheral relative to central populations in *Picea sitchensis* (Pinaceae). *Am J Bot* 94:991–998
- Mitton JB (1992) The dynamic mating system of conifers. *New For* 6:197–216
- Morgan MT, Wilson WG (2005) Self-fertilization and the escape from pollen limitation in variable pollination environments. *Evolution* 59:1143–1148
- Morgante M, Vendramin GG, Rossi P (1991) Effects of stand density on outcrossing rate in two Norway spruce (*Picea abies*) populations. *Can J Bot* 69:2704–2708
- Myczkowski S, Bednarz Z (1974) The stone pine *Pinus cembra* L. In: S. Myczkowski (ed.), *Rodzime drzewa Tatr I. Studia Ośrodka Dokumentacji Fizjologicznej* 3: 111–139 (in Polish with English summary)
- Neal DB, Adams WT (1985) Allozyme and mating-system variation in balsam fir (*Abies balsamea*) across a continuous elevational transect. *Can J Bot* 63:2448–2453
- O’Connel LM, Mosseler A, Rajora OP (2006) Impacts of forest fragmentation on the mating system and genetic diversity of white spruce (*Picea glauca*) at the landscape level. *Heredity* 97:418–426
- O’Connel LM, Mosseler A, Rajora OP (2007) Extensive long-distance pollen dispersal in a fragmented landscape maintains genetic diversity in white spruce. *J Hered* 98:640–645
- Paryski W (1971) Sadzenie i przesadzanie limby. In: S Białobok (Ed) *Limba Pinus cembra* L. Nasze drzewa leśne 2:50–56 (in Polish)
- Politov DV, Krutovskii KV (1994) Allozyme polymorphism, heterozygosity, and mating system of stone pines (*Pinus*, subsection *Cembrae*). In: Schmidt WC, Holtmeier F-K (eds) Proceedings of “international workshop on subalpine stone pines and their environment: the status of our knowledge”. US Department of Agriculture Forest Service, Intermountain Research Station, Ogden, Utah, pp 36–42
- Politov DV, Pirko YV, Pirko NN, Mudrik EA, Korshikov II (2008) Analysis of mating system in two *Pinus cembra* L. populations of the Ukrainian Carpathians. *Ann ForRes* 51:11–18
- Porcher E, Lande R (2005) The evolution of self-fertilization and inbreeding depression under pollen discounting and pollen limitation. *J Evol Biol* 18:497–508
- Rajora OP, Mosseler A, Major JE (2002) Mating system and reproductive fitness traits p feaster white pine (*Pinus strobus*) in large, central versus small, isolated, marginal populations. *Can J Bot* 80:1173–1184
- Restoux G, Silva DE, Sagnard F, Torre F, Klein E, Fady B (2008) Life at the margin: the mating system of Mediterranean conifers. *Web Ecol* 8:94–102
- Ritland K (1990) Inferences about inbreeding depression based on changes of the inbreeding coefficient. *Evolution* 44:1230–1241
- Ritland K (2002) Extensions of models for the estimation of mating systems using n independent loci. *Heredity* 88:221–228
- Robledo-Arnuncio JJ, Smouse PE, Gil L, Alía R (2004) Pollen movement under alternative silvicultural practices in native populations of Scots pine (*Pinus sylvestris* L.) in Central Spain. *Forest Ecol Manag* 197:245–255
- Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2006) A new method of estimating the pollen dispersal curve independently of effective density. *Genetics* 173:1033–1045
- Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2007) POLDISP: a software package for indirect estimation of contemporary pollen dispersal. *Mol Ecol Notes* 7:763–766
- Salzer K (2011) Wind- and bird-mediated gene flow in *Pinus cembra*: Effects on spatial genetic structure and potential close-relative inbreeding. The PhD Thesis, University of Zurich, Faculty of Science
- Salzer K, Gugerli F (2012) Reduced fitness at early life stages in peripheral versus core populations of Swiss stone pine (*Pinus cembra*) is not reflected by levels of inbreeding in seed families. *Alp Bot* 122:75–85
- Salzer K, Sebastiani F, Gugerli F, Buonamici A, Vendramin GG (2009) Isolation and characterization of polymorphic nuclear microsatellite loci in *Pinus cembra* L. *Mol Ecol Resour* 9:858–861
- Savolainen O, Hedrick P (1995) Heterozygosity and fitness: no association in Scots pine. *Genetics* 140:755–766
- Savolainen O, Kärkkäinen K, Kuitinen H (1992) Estimating numbers of embryonic lethals in conifers. *Heredity* 69:308–314
- Schwendemann AB, Wang G, Mertz ML, McWilliams RT, Thatcher SL, Osborn JM (2007) Aerodynamics of saccate pollen and its implications for wind pollination. *Am J Bot* 94:1371–1381
- Shea KL (1987) Effects of population structure and cone production on outcrossing rates in Engelmann spruce and subalpine fir. *Evolution* 41:124–136
- Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* 55:260–271

- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. *J R Statist Soc B* 64:583–639
- Takayama S, Isoagi A (2005) Self-incompatibility in plants. *Annu Rev Plant Biol* 56:467–489
- Tomback DF, Holtmeier FK, Mattes H, Carsey KS, Powell ML (1993) Tree clusters and growth form distribution in *Pinus cembra*, a bird-dispersed pine. *Arct Alp Res* 25:374–381
- Ulber M, Gugerli F, Bozi G (2004) Technical guidelines for genetic conservation and use of Swiss stone pine (*Pinus cembra* L). Tech Rep International Plant Genetic Resources Institute
- Vogler DW, Kalisz S (2001) Sex among the flowers: the distribution of plant mating systems. *Evolution* 55:202–204
- Wright S (1931) The evolution in Mendelian populations. *Genetics* 16: 97–159
- Wright SI, Ness RW, Foxe JP, Barret SCH (2008) Genomic consequences of outcrossing and selfing in plants. *Int J Plant Sci* 169:105–118
- Young AG, Broadhurst LM, Thrall PH (2012) Non-additive effects of pollen limitation and self-incompatibility reduce plant reproductive success and population viability. *Ann Bot* 109:643–653
- Zwijacz-Kozica T, Żywiec M (2007) Fifty-year changes in a strictly protected stone pine population in the Tatra National Park. *Nat Conserv* 64:73–82