Genetic structure of a regionally endangered orchid, the dark red helleborine (*Epipactis atrorubens*) at the edge of its distribution

Hilde Hens[1,2] *, Anne Jäkäläniemi[3], Kadri Tali[4], Petr Efimov[5], Alexey V. Kravchenko[6], Laura Kvist[1]

[1] Department of Ecology, University of Oulu, P.O. Box 3000, 90014 Oulu, Finland

[2] Thule Institute, University of Oulu, P.O. Box 7300, 90014 Oulu, Finland

[3] Thule Institute, Oulanka Research Station, Liikasenvaarantie 134, 93999 Kuusamo, Finland

[4] Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Kreutzwaldi 5, Tartu 51014, Estonia

[5] Komarov Botanical Institute of the Russian Academy of Sciences, Prof. Popov str. 2, 197376 St. Petersburg, Russia

[6] Karelian Research Centre of the Russian Academy of Science, Pushkinskaya St. 11, 185910 Petrozavodsk, Russia

Corresponding author

Hilde Hens

Department of Biology, University of Oulu, P.O. Box 3000, 90014 Oulu, Finland

hilde.hens@oulu.fi

Acknowledgments

Abstract

The genetic structure and diversity of species is determined by both current population dynamics and historical processes. Population genetic structure at the edge of the distribution are often expected to differ substantially from populations at the centre, as these edge populations are often small and fragmented. In addition, populations located in regions that have experienced repeated glaciations throughout the Pleistocene, may still carry imprints from the genetic consequences of frequent distribution shifts. Using chloroplast DNA sequences and nuclear microsatellite markers we studied the genetic structure of *Epipactis atrorubens* at the northern edge of its distribution. Contrary to populations in the centre of the distribution, populations at the northern range are regionally endangered as they are small and disjunct. Sequence data of two chloroplast loci and allelic data from six nuclear microsatellite markers were obtained from 297 samples from Finland, Estonia and Russia. We sought for genetic indicators of past population processes, such as post-glacial colonisation history of *E. atrorubens*. As expected, we observed low genetic variation, in terms of numbers of substitutions, haplotypes and alleles, and significant levels of differentiation, especially pronounced in the chloroplast DNA. These features suggest that the edge populations could be prone to extinction.

Introduction

The population density is assumed to decline from the centre towards the edge of the species distribution (abundant-center hypothesis; Brown 1984; Brussard 1984; Lawton 1993; Sagarin and Gaines 2002). The genetic diversity of populations at the edge of the distribution is, therefore, often assumed to be low compared to populations at the centre (Brussard 1984; Eckert et al. 2008). In addition, populations at the edge of the distribution often experience unfavourable environmental conditions compared to populations in the central part of the distribution where the conditions are often close to optimal, resulting in smaller population sizes and consequently lower genetic variation at the edge (Brussard 1984; Sagarin and Gaines 2002).

Further, populations especially at the northern edge of the distributions can have lower genetic variation than elsewhere as a consequence of their glacial history. Boreal species have experienced repeated glaciations throughout the Pleistocene. The coverage of north-western Europe by large continental ice-sheets, with surrounding areas of permafrost, low temperatures and reduced water availability has caused repeated distributional shifts of boreal and temperate species (Hewitt 1999, 2000, 2004). These changes have contributed to the present geographic distribution and genetic structure of these species (Hewitt 1999, 2004). Populations restricted to refugia were often isolated due to geographical barriers and experienced genetic bottlenecks. During inter-glacial periods, range expansions were commonly accompanied by repeated founder events, resulting in a loss of genetic variation. Therefore, northern populations are often assumed to have lower levels of genetic variation, compared to species closer to the refugial areas.

However, the demographic effects of these repeated range contractions and expansions depended on the life-history parameters of the species, such as dispersal capacity, reproductive rates, habitat requirements and interactions with other species (Taberlet et al. 1998; Hewitt 1999). The pace of colonization also matters, as rapid colonization results more likely in genetic homogeneity among populations than slow colonization. During a slow colonisation, genetic drift, bottlenecks and local adaptations can have larger effects on populations, leading to an increased differentiation among populations compared to a fast colonisation (e.g. Knutsen et al. 2013). Thus, peripheral populations along the northern edge of the species distribution are assumed to consist of genetically

differentiated small populations harbouring low levels of genetic diversity, due to the combined effects of random genetic drift, reduced gene flow, increased inbreeding and founder effects (Lesica and Allendorf 1995; Durka 1999; Hutchison and Templeton 1999; Lammi et al. 1999; Eckstein et al. 2006; Pfeifer et al. 2009). Knowledge of present population genetic patterns and of the historical processes that have lead to these patterns is crucial to understand forces that affect population viability and at the same time gives tools for designing effective conservation strategies for endangered species. This knowledge is becoming more and more urgent for northern populations, as climate change is projected to have a larger impact in northern than in southern regions.

Orchid species are often assumed to harbour high levels of genetic diversity and low levels of population genetic structuring, as many species are outcrossing, long-living and produce small wind dispersed seeds with a high potential of effective dispersal (Swarts and Dixon 2009). However, the genetic structure and diversity has been documented to vary extensively among species and within species along the distribution of orchid species (e.g. Gustafsson 2000) and studies have documented also remarkably low levels of genetic variation in several orchid species (e.g. Chung and Chung 2007; Chung 2009). Many orchids have wide distributions, covering different types of habitats and environmental conditions.

In this study, we investigated the spatial genetic structure of *Epipactis atrorubens* populations at the northern edge of its geographical distribution. Whereas the species is common in the central part of its distribution, populations are small and fragmented at the northern edge of the range. To quantify the population genetic structure and the genetic variation of the populations, two chloroplast loci were sequenced and six nuclear microsatellite markers were used. In addition, we discuss the historical processes underlying the present distribution and genetic structure. As the studied populations are small and isolated and located at the edge of the distribution, we expected low levels of genetic diversity within the populations and high levels of genetic differentiation between populations.

Material and Methods

Study species

The dark red helleborine, *Epipactis atrorubens* (Hoffm. Ex Bernh.) Besser, is a perennial orchid species, which is widely distributed across Europe and extending to the Ural Mountains in Asia and western Siberia (Fig. 1; Tuulik 1998; Efimov 2004). Populations mainly occur in Central and Eastern Europe, with northern populations found in the south of Sweden, along the coast line of Norway and in Finland (Meusel et al. 1992; Anderberg 2001), but at the northern edge of the distribution the population are very disjunct. The habitat consists mainly of slopes and shorelines and is characterized by limestone or dolomite. According to the IUCN red list criteria, in Europe, *E. atrorubens* is currently not categorised as endangered (Least concern; Bilz et al. 2011). However, regionally the species has been categorized as endangered at the northern edge of its distribution. *E. atrorubens* is considered vulnerable in Finland (Rassi et al. 2010) and in the adjacent regions in Russia it is listed as endangered (Murmanskaya; Konstantinova et al. 2014) or vulnerable (Leningradskaya, Vologodskaya, Karelia; Tzvelev 2000; Konechnaya and Suslova 2004; Ivanter and Kuznetsov 2007). In Estonia, Norway and Sweden the species is categorized as least concern (Lilleleht 2008; Kålås et al. 2010; Westling 2015). The life history of *E. atrorubens* is characterized by a predominately allogamous reproduction (Brzosko et al. 2006; Talalaj and Brzosko 2008; Jakubska-Busse and Kadej 2011), it is pollinated by wasps and bumblebees (Brzosko et al. 2006; Tałałaj and Brzosko 2008), has a protocorm life stage, with a close association with mycorrhiza and overwinters as an underground rhizome and has a prolonged dormancy (Jäkälaniemi et al. 2011).

Sampling

In total, 297 samples were collected during 2011-2013 in three regions in Finland, covering the whole Finnish distribution, one region in Estonia and three regions in Russia (Fig. 1, Table S1). One leaf was sampled per individual and frozen at -20°C or dried with silica gel. As one individual can have several shoots (1-28 shoots; unpubl. data) that share a single rhizome, only one leaf was sampled from a group of shoots.

Molecular methods

Total DNA was extracted using a modified 2x cetyltrimethylammonium bromide (CTAB) protocol (Doyle 1987). Five chloroplast loci were tested for sequencing. Three of these loci were excluded due to the low amplification success (trnS-trnG intergenic spacer and trnQ-rps16 intergenic spacer; Shaw et al. 2007) or the absence of genetic variation (trnL intron; Taberlet et al. 1991). Two chloroplast loci, trnK^{uuu} intron (Johnson and Soltis 1994) and intergenic spacer of the rbcL-accD region (Tranchida-Lombardo et al. 2011), were sequenced in this study. In addition, six nuclear microsatellite loci, Ccal25 (Minasiewicz and Znaniecka 2014), SW1-76, SW1-78, SW1-120, SW2-152 and SW2-185 (Tranchida-Lombardo et al. 2008), were selected to investigate nuclear genetic variation and differentiation. The amplification conditions, are given in the supplementary material (Appendix 1 and table S2). The chloroplast sequences and nuclear microsatellites were run with an ABI PRISM 3730 automatic sequencer (Applied Biosystems) and microsatellite alleles scored as allele sizes with GeneMapper v.5 (Applied Biosystems) (see Appendix 1).

Statistical analyses

Genetic diversity

In Russian Karelia, one individual, reported as being possible *Epipactis helleborine* (L.) Crantz based on the morphology, had a unique haplotype in both chloroplast loci. In a study by Tranchida-Lombardo et al. (2011), this haplotype was found only in *E. helleborine* and therefore we excluded it from further analyses. The chloroplast sequences were aligned manually using BioEdit 7.0.9.0 (Hall 1999). Haplotype diversities (Hd; Nei 1987) and the Watterson's estimators of theta (Θ_S; Tajima 1989) were calculated with DnaSP 5.10 (Librado and Rozas 2009). The nucleotide diversities (π; Nei 1987) and haplotype frequencies were calculated with Arlequin v. 3.5 (Excoffier and Lischer 2010). A parsimony network of the haplotypes was drawn using TCS 1.21 (Clement et al. 2000). The two chloroplast loci were combined for all analyses.

In order to quantify genotyping errors due to scoring errors and allelic dropouts, the mean error rate per allele and per locus (Pompanon et al. 2005) was calculated for the nuclear microsatellites, by repeating the genotyping of randomly (>5% of the sample size) chosen samples. In addition, the data was checked with Micro-

checker (Van Oosterhout et al. 2004). GENEPOP 4 (Raymond and Rousset 1995; Rousset 2008) was used to test for deviations from linkage equilibrium within the sampled populations. The allele frequencies, allelic richness (A_R corrected for sample size using the rarefaction method) and the inbreeding coefficients (F_{IS}) were calculated with FSTAT v2.9.3.2 (Goudet 2001) and the significance of heterozygote excess or deficit was tested using 10 000 permutations. The observed (H_O) and expected (H_E) heterozygosities were calculated using a resampling with replacement (1000 bootstrap replicates) for each populations using the PopGnKit package v1 (Paquette 2012) in R 3.2.3.

Genetic differentiation

An allele size permutation test (Hardy et al. 2003) was used to test whether allele size based statistics are more adequate than allelic state based statistics (i.e. stepwise mutation model versus infinite allele model). The test was run in SpaGeDi 1.5 (Hardy and Vekemans 2002) using 10 000 permutations.

To estimate the differentiation among regions, analyses of molecular variance (AMOVA; Excoffier et al. 1992) (10 000 permutations) were conducted. Pairwise and Φ_{ST}-values (chloroplast loci) and F_{ST}-values (nuclear loci) were calculated (10 000 permutations) with Arlequin v. 3.5 (Excoffier and Lischer 2010) for the chloroplast loci and the nuclear loci separately. No substitution model was used for estimation of Φ_{ST}-values and AMOVA, since both loci had only one substitution. In addition to the pairwise F_{ST}-values, pairwise Jost's differentiation D values (D_{est}; Jost 2008) were calculated using SMOGD 1.2.5 (Crawford 2010) for the nuclear markers, as D_{est} values are independent of the within-subpopulation heterozygosities and nearly unbiased by sample size.

The nuclear genetic structure was invested with STRUCTURE 2.3 (Pritchard et al. 2000; Falush et al. 2003). Correlated allele frequencies between populations and an admixture model was assumed and the length of the burn-in period was 100 000 with 500 000 MCMC repeats after the burn-in. We tested for $1 \leq K \leq 10$ (K= number of clusters) and did 10 replicates for each K. ΔK was calculated with the Evanno method (Evanno et al. 2005) using STRUCTURE Harvester 0.6.94 (Earl and vonHoldt 2012). CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007) was used to align the membership coefficients of the repeated runs of STRUCTURE and Distruct 1.1 (Rosenberg 2004) to construct the bar plots of the individual assignment probabilities to the clusters.

The spatial genetic structure was investigated using a SAMOVA (Spatial Analyses of Molecular Variance; Dupanloup et al. 2002) to make inferences about the spatial genetic structure. In this analysis, groups of populations that are maximally differentiated from each other are defined by maximizing the proportion of genetic variance among groups (F_{CT}), so that the groups are geographically homogeneous. The SAMOVA was performed for $2 \leq K \leq 4$ (K = number of groups) with 100 simulated annealing processes, for the chloroplast and the nuclear markers separately. Spatial genetic structuring was further investigated by testing for isolation by distance (IBD). Tests for isolation by distance were conducted on a population level and an individual level. On the population level, the correlation between the genetic distances (Slatkin's linearized F_{ST}; Slatkin 1995) and the logarithm of geographical distances of the sampling sites was calculated by conducting a Mantel test (Mantel 1967) using 10 000 permutations in Arlequin v. 3.5 (Excoffier and Lischer 2010). This was done separately for the chloroplast sequence data and nuclear microsatellite data. The isolation by distance (IBD) was tested on an individual level using the nuclear markers. The correlation between the mean pairwise kinship coefficient (Fij) (Loiselle et al. 1995) and the natural logarithm of geographical distances between the populations was tested in SPaGeDi 1.5 (Hardy and Vekemans 2002) using nine distance classes and 1 000 permutations. Standard errors were estimated using a jack-knife method over loci and significance using a one-tailed permutation test (10 000 permutations).

Past population size changes

In order to infer the demographic history of populations using the chloroplast data, different indicators of past populations size changes were calculated in DnaSP 5.10 (Librado and Rozas 2009). Because the program DnaSP does not incorporate gaps in the aligned files in the analyses, gaps as a result of the poly(A) repeat in the trnK[uuu] locus were replaced by a nucleotide (G), so that insertions and deletions were also included as variation. Consequently, the weight of substitutions and insertions/deletions was set as equal. In order to test for population growth, Tajima's D (Tajima 1989), Fu's Fs (Fu 1997), raggedness index (r, Harpending 1994) and Ramos-Onsins and Rozas's R^2 statistics (Ramos-Onsins and Rozas 2002) were calculated. P-values were obtained by a coalescence test with 10 000 replicates, significance set to $p < 0.05$ (Fu's Fs, $p < 0.02$; Excoffier and Lischer 2010). The timing of potential past population size changes were estimated by the calculation of τ (Rogers and Harpending

1992). In order to make inferences about past population size changes using the nuclear loci, a Wilcoxon rank test and a mode shift test were conducted in Bottleneck 1.2.002 (Piry et al. 1999) using the two-phase mutation model (TPM) with 70 % single step mutations. The Garza Williamson Index that informs of gaps in the allele size distribution (G-W; Garza and Williamson 2001) was calculated with Arlequin v. 3.5 (Excoffier and Lischer 2010). An index lower than 0.68 is proposed as evidence for a bottleneck (Garza and Williamson 2001).

Results

Genetic diversity

When combining the chloroplast sequence loci, a 716 bp sequence (468 bp of rbcL-accD, 544 bp trnK^{uuu}) was obtained from 226 samples. In total, seven haplotypes were identified (rbcL-accD: 2 haplotypes, trnK^{uuu} locus: 5 haplotypes). The haplotypes differed from each other by two substitutions (one in the rbcL-accD locus and one in the trnK^{uuu} locus) and one poly(A) repeat with nine, ten or eleven nucleotides (trnK^{uuu}). The sequences have been submitted to GenBank (Accession numbers: rbcL-accD KM671282 – KM671528, trnK^{uuu} KM671529 – KM671782).

The haplotype frequencies and the genetic variation varied among the regions (Table 1, Table S3, Table S4, Fig. 2). The most common haplotype was haplotype CA10, which was found in 63.7 % of the individuals (Table S4, Fig. 2). This haplotype was observed in all populations and was the only haplotype in populations Juankoski and Hanko. It was also the most common haplotype in all populations, except populations Vologda and Karelia. All seven haplotypes were found in Oulanka. No more than two haplotypes were observed in Karelia, differing only in the trnK^{uuu} locus. In the three remaining populations, three haplotypes were observed in each.

The mean error rate of the nuclear microsatellite markers was 0.034 per locus. Indications for null alleles were found as an excess of homozygotes in different populations for different loci (SW2-185 Oulanka; SW2-152 Juankoski; SW1-76 Oulanka, Juankoski; SW1-78 Oulanka, Estonia, Hanko; Ccal25 Oulanka, Juankoski). Indications for scoring errors due to stuttering bands were observed in two loci (SW2-185 Oulanka; SW1-78 Oulanka, Estonia, Hanko). Due to the absence of indications of null-alleles or stuttering bands in other populations,

we used all the loci in further analyses. No significant linkage disequilibrium was detected between the loci. The allele frequencies and genetic diversity estimates are given in Tables S5 and 2, respectively. All loci had low levels of polymorphism with only two (SW1-76), three (Ccal25, SW1-120, SW2-152), or four (SW1-78, SW2-185) alleles per locus. Private alleles were observed only in Oulanka (SW1-120: allele 88 N=6; SW1-78: allele 236 N=5) and Hanko (SW2-185: allele 278 N=9). Deviations from the Hardy-Weinberg equilibrium were found in two populations as a deficit of heterozygotes and in one population as an excess of heterozygotes (Table 2).

Genetic differentiation

The allele size permutation test indicated that stepwise mutations do not contribute significantly to the genetic differentiation (p=0.876). Therefore F-statistics and kinship coefficient based on allelic state were used in further analyses of the microsatellite loci. The AMOVA analyses showed significant overall differentiation values of Φ_{ST}=0.236 (p < 0.001) and F_{ST}=0.071 (p < 0.001) for the combined chloroplast loci and the nuclear loci, respectively (Table S6). The pairwise Φ_{ST}-values, F_{ST}-values and D_{est} are given in Table 3 and Table S7. As a significant correlation was found between the pairwise F_{ST} and D_{est} values (p<0.023, r2= 0.495; tested by using a Pearson's product-moment correlation in R), F_{ST} values are used for further discussion. Out of the 21 pairwise differentiation values (Φ_{ST} and F_{ST}) 12 and 16 were significant, for the nuclear and chloroplast data respectively, whereof 8 differentiation values were significant for both the nuclear and chloroplast data.

The Bayesian structuring analysis show that the best structuring occurs at K=2, indicated by the highest log probability (ln P(D)) and the highest ΔK (Fig. 3). A second peak for ΔK was observed for K=4. The assignment probabilities for each individual are given in Fig. 3 for K=2 and K=4. Individuals from all sampled regions were assigned to both clusters with varying probabilities, indicating admixture. However, a higher proportion of individuals assigned to cluster 2 can be observed in populations Vologda and Pskov.

A SAMOVA was conducted in order to find the best groupings among the sampled regions (Table S8). For the nuclear loci, in K=2 the among-group variation explained a high percentage (33.60%) of the total variation. The among population variation within groups was low (-12.66%; i.e. zero). In this grouping, the population Hanko was identified as a different group from all the other populations. For K=4, the among group variation was also

high (35.59%). In these groupings, Hanko was identified as a separate group, Oulanka and Karelia was another group, Pskov and Vologda was a third group and Juankoski and Estonia as a fourth group. For the chloroplast loci, the variation explained by the groupings increased with the number of groups, whereas the variation due to variation among populations within groups decreased, indicating strong differentiation among the regions. Juankoski, Hanko and Pskov were combined as one group in all three analyses. When K was increased up to 6, Juankoski and Hanko continued to be placed together (data not shown), while other populations formed one group each, indicating that the two populations are closely related to each other.

The Mantel test revealed no correlation between the genetic and geographical distances among regions for neither the combined chloroplast loci (r = -0.186, p = 0.816) nor the nuclear loci (r =0.089, p = 0.331), indicating that the observed patterns of genetic differentiation are not simply explained by isolation by distance. In addition, no signs of isolation by distance were found for the nuclear loci on individual level, as no significant negative correlation of the mean kinship coefficient and the natural logarithm of the geographic distance was observed (r2=<0.001, p=0.267).

Demographic history

No indications for past population size changes were detected with the chloroplast data (Table S9). However, indications for population size changes were found in all populations using the nuclear microsatellites (Table 2). The mode shift test indicated shifted allele frequency distributions caused by population size reductions in all populations. The Garza-Williamson index was low in all populations (0.397-0.556) and the Wilcoxon test showed a significant excess in the observed heterozygosity compared to the expected heterozygosity based on the number of alleles, further indicating a bottleneck. The absence of signs of population size changes in the chloroplast loci might be explained by the low genetic variation, as the number of variable sites might have been insufficient to provide enough resolution for making inferences about the demographic history.

Discussion

Genetic diversity

We found low levels of genetic variation within the studied regions in terms of number of chloroplast haplotypes (7), variable nucleotide sites (4) and the number of microsatellite alleles. In addition, populations of the different regions were differentiated both in the nuclear and chloroplast DNA. The maximum parsimony network shows that the detected chloroplast haplotypes in this study are only one to three mutational steps from the most common haplotype (Fig. 2). The heterozygosity was moderate-low (mean H_E=0.256; 0.123-0.371) in the regions. Genetic diversity in orchids has been found to vary much. Higher levels of genetic variation have been reported in several orchid species that were studied using nuclear microsatellite markers (H_E= 0.77-0.92 for *Ophrys sp.* (Soliva and Widmer 2003); H_E=0.62-0.81 for *Gymnadenia conopsea* (Gustafsson 2000); H_E= 0.355–0.544 for *Cypripedium kentuckiense* (Pandey et al. 2015)). Such high levels of genetic diversity are generally expected in perennial outbreeding orchids with wind-mediated seed dispersal (Swarts and Dixon 2009; Templeton and Levin 1979; Jersáková and Malinová 2007). However, similarly to our results, low levels of genetic diversity have been documented as well in several orchid species (H_E=0.000-0.065 for *Cephalanthera longibracteata* (Chung et al. 2004); H_E=0.191-0.322 for *E. helleborine* (Ehlers and Pedersen 2000)). Using allozyme markers, a low genetic variation, or even a lack of genetic variation, has been documented in *Epipactis* species before, including *E. atrorubens* (H_E=0.000 for *E. papillosa* (Chung 2009);H_E=0.006-0.019 for *E. thunbergii* (Chung and Chung 2007); H_E=0.000-0.204 for *E. helleborine* (Hollingsworth and Dickson 1997); H_E=0.025-0.052 *E. atrorubens* (Brzosko et al. 2006)). In addition, low plastid genetic variation was found in Italian *E. atrorubens* populations (H_E=0.000; Tranchida-Lombardo et al. 2011).

The levels of genetic diversity are determined by both current local population dynamics, adaptations to local conditions and historical processes. Low levels of genetic diversity in orchids are frequently attributed to the effects of inbreeding and genetic drift as populations are often small and fragmented and gene flow limited. Genetic drift forms likely an important causal factor for the low levels of genetic variation in the studied populations, as in addition of being small and fragmented, they are located at the edge of a more continuous distribution in central

Europe. However, it is worth mentioning that low levels of genetic variation have been found for *E. atrorubens* populations in other regions (i.e. Poland (Brzosko et al. 2006) and Italy (Tranchida-Lombardo et al. 2011)). The low genetic variation of the *E. atrorubens* populations could also be related with infrequent reproduction and seedling emergence of *E. atrorubens*, as reported by Brzosko et al. (2006). Indeed, low seedling recruitment rates has been detected in *E. atrorubens* populations in the Oulanka populations (unpublished data) and in the Polish populations (Brzosko et al. 2006).

The low observed genetic variation in the studied populations was especially pronounced in the populations Hanko and Juankoski, where no chloroplast diversity was observed. In addition, the population Karelia only had two haplotypes for the combined chloroplast loci, which differed only in one insertion/deletion. The Hanko population was first observed in 2003 and grows in a road verge. Populations of *E. atrorubens* are frequently observed in human-influenced habitats, such as roadsides, also in regions where limestone has been transferred and thereby new suitable habitats for the species have been created (Tuulik 1998; Brzosko et al. 2006). These types of populations can suffer from the negative genetic consequences of founder events, as the populations are likely established by a small number of seeds. The low levels of genetic variation in Juankoski and Karelia are more likely caused by random genetic drift than by founder events because populations in these regions are older. In addition, the Juankoski region harbours only a few patches of subpopulations (6) which are very small and isolated from other parts of the distribution.

Genetic structure

The loss of genetic variation is expected to be counteracted by gene flow. However, population genetic structuring among the *E. atrorubens* populations is observed, indicating limited gene flow. This was suggested by varying allele and haplotype frequencies between the regions, pairwise ϕ_{ST}- and F_{ST}- estimates, the AMOVA- and the SAMOVA-analyses. The disjunct distribution of *E. atrorubens* likely prevents gene flow between the regions, which contributes to the genetic differentiation of the populations. Restricted gene flow for this species has been reported also by Brzosko et al. (2006) and on a fine spatial scale in the Oulanka region (unpublished data). The chloroplast loci showed more differentiation than the nuclear loci, which was expected due to the smaller effective

population size of chloroplast genes (Birky et al. 1989), resulting in larger impacts of genetic drift in chloroplast DNA compared to nuclear DNA (Petit et al. 1993, 2005). In addition, differences in seed and pollen flow can explain different genetic differentiation patterns between chloroplast and nuclear markers. Gene flow of chloroplast DNA occurs through seed dispersal and nuclear gene flow through pollen and seed dispersal (Petit et al. 1993). The chloroplast differentiation (Φ_{ST}=0.236) indicates restricted seed dispersal. However, due to their small and light seeds, orchids are expected to have the ability for long-distance seed dispersal by wind (Murren and Ellison 1998; Arditti and Ghani 2000; Jersáková and Malinová 2007) and such long distance seed migration has been documented in European orchids (e.g. Vanden Broeck et al. 2014). Yet, also restricted seed dispersal has been shown in several orchid species (Murren and Ellison 1998; Machon et al. 2003; Chung et al. 2004) as well as in *E. atrorubens* (unpublished data). The long-distance seed dispersal, even when it occurs, may not be effective due to a low germination success and seedling recruitment, as orchids are highly dependent on associations with mycorrhiza (Davies et al. 1983; Rasmussen 1995) and *E. atrorubens* has been shown to have strong mycorrhizal fungal specialisation during the protocorm stage (Bidartondo and Read 2008). A higher germination success in the proximity of adult individuals, due to both a higher abundance of mycorrhiza close to the adult individuals and abiotic factors, has been shown in different orchid species (Batty et al. 2001; Diez 2007). Thus, low effective seed dispersal can result from a low germination success, rather than an inability of long-distance seed dispersal.

The nuclear differentiation on the other hand was low, indicated by the low differentiation value (F_{ST}=0,071) and the Bayesian structure analyses. The Bayesian structure analyses indicated the presence of two genetic clusters and potential sub-structuring within the two clusters. However, plotting the assignment probabilities of each individual highlights that there is still a high admixture among the isolated populations. In addition to varying levels of divergence between the nuclear and chloroplast loci, the patterns differed between the markers indicating that the seed and pollen flux among the regions is different.

The nuclear SAMOVA analyses indicate that the Hanko population is differentiated from the other regions. This likely results from a recent colonisation of the population. The geographically closest population in Estonia was assumed to be the source for the Hanko population. However, the high frequency of a private allele (0.500; allele 278 SW2-185) found in Hanko indicates that the population was likely founded by an unsampled population.

From a conservation point of view, this kind of recent establishments of new populations due to the anthropogenic development of novel suitable habitats, as observed in Hanko, can increase the connectivity of disjunct populations in the future.

Historical processes

The genetic consequences of the long post-glacial colonisation history of species can last for a long time in species with a long generation time, such as *E. atrorubens* (Stenström et al. 2001). The sampled regions have been ice-free for 11 000 and 14 400 years (Svendsen et al. 1999; Tikkanen and Oksanen 2002; Kallio 2006). The observed differentiation patterns can indeed reflect the colonisation history of these regions after the LGM and, consequently, the different successive ages of populations (Stenström et al. 2001). European regions covered by continental ice or permafrost during the LGM are generally considered to be colonized after the LGM from the Iberian and Balkan refugia (Taberlet et al. 1998). However, species with a wide distribution are expected to have occupied multiple refugia during the LGM (Steward et al. 2010). The presence of private nuclear alleles in Oulanka and the presence of the chloroplast haplotype A11 (the trnK^{UUU} locus) only in the northern regions Karelia and Oulanka, could indicate that these regions were colonized from different refugia that were not involved in colonization of southern populations. Indeed, Scandinavia has been shown to be a contact zone for multiple colonization routes in different species (Taberlet et al. 1995, 1998; Jaarola and Tegelström 1995). A colonisation pathway from the North, via Norway or from the East could have led to founding of the northern populations, Oulanka and Karelia.

After the LGM, species expanded their range following the retreat of the ice, leading to repeated founder events. Such repeated founder events can result in a loss of alleles and an increased homozygosity, resulting in lower genetic variation of populations in areas previously covered with continental ice, compared to populations located in the refugia (Hewitt 1996; Provan and Bennet 2008). While a reduction in the genetic variation is expected with a fast colonisation, more genetic variation is maintained with a gradual or slow expansion. Our findings are more consistent with a slow expansion, as even though we observed quite low levels of genetic variation, there were no signs of reduced variation towards the north or isolation-by-distance, which would be

expected under a fast expansion. However, additional historical processes could have led to loss of genetic variation among the southern populations. If the dispersal rates reflect colonisation rates, low colonisation rates could be a result from low seed dispersal levels. In addition, other life history characteristics, such as the reproduction mode and growth form can affect colonisation rates. *E. atrorubens* is a perennial species with a long generation time, which can slow down the colonization pace. Further, it is pollinated by wasps and bees and animal pollinated species are considered to have slower colonisation rates compared to wind pollinated species (Govindaraju 1988).

Conservation implications

Climate change can be a major threat for the persistence of species by changing the habitat conditions, especially for species with very strict habitat conditions, such as *E. atrorubens*. Populations can cope with environmental changes by different means. Species can avoid the negative consequences of changed environmental conditions by adapting to the new conditions. However, low levels of genetic variation, as observed for this species, can elude the possibility of species to adapt to their new environment. Gene flow in small disjunct populations may be insufficient to counteract the loss of genetic diversity due to genetic drift or limited number of founders. We also detected recent population size reductions with the nuclear markers. Therefore, the disjunct distribution, associated with their specific habitat requirements, low levels of gene flow, signs of population size decreases and limited standing genetic variation could make the populations vulnerable for future environmental changes, such as climate change, since they might not have enough evolutionary potential to adapt to new environments. Alternatively, species can shift their ranges towards more suitable habitats. Yet, low dispersal rates, as indicated by the differentiation patterns, could imply that the range shift in response to environmental changes is not fast and effective enough. Consequently, low dispersal rates coupled with the low genetic variation and specific habitat requirements, might have a strong negative impact for the species viability, as it may not be able to keep up with the fast climate change. Management of the populations of *E. atrorubens* to avoid further population declines and preventing further loss of genetic variation is needed to avoid extinctions of these populations.

References

Anderberg AA (2001) *Epipactis atrorubens*. Naturhistoriska riksmuseet.
http://linnaeus.nrm.se/flora/mono/orchida/epipa/epipatr.html. Accessed 23 September 2015

Anthos (2010) Information System of the plants of Spain. Real Jardín Botánico, CSIC. Fundación Biodiversidad.
http://www.anthos.es. Accessed on 8 August 2016

Arditti J, Ghani AKA (2000) Tansley Review No. 110. Numerical and Physical Properties of Orchid Seeds and
Their Biological Implications. New Phytol 145: 367-421

ArtDatabanken (2016) The Swedish Species Information Centre. http://artfakta.artdatabanken.se. Accessed on 8
August 2016

Batty AL, Dixon KW, Brundrett M, Sivasithamparam K (2001) Constraints to symbiotic germination of terrestrial
orchid seed in a mediterranean bushland. New Phytol 152:511-520

BfN (2016) FloraWeb Daten und Informationen zu Wildpflanzen und zur Vegetation Deutschlands. URL:
http://www.floraweb.de. Cited 08.06.2016 (in German)

Bidartondo MI, Read DJ (2008) Fungal specificity bottlenecks during orchid germination and development. Mol
Ecol 17: 3707–3716

Bilz M, Kell SP, Maxted N, Lansdown RV (2011) European Red List of Vascular Plants. Publications Office of
the European Union, Luxembourg

Birky CW, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration, mutation, and drift:
equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear
genes. Genetics 121:613-627

Brown JH (1984) On the relationship between abundance and distribution of species. Am Nat 124: 255-279

Brussard PF (1984) Geographic patterns and environmental gradients: the central-marginal model in Drosophila
revisited. Annu Rev Ecol Syst 15: 25-64.

Brzosko E, Tałałaj I, Wróblewska A (2006) Genetic structure of rare *Epipactis atrorubens* populations from two
national parks in Northeast Poland. Polish Botanical Studies 22:71-80

BSBI, BRC (2008) Online atlas of the British & Irish Flora. http://www.brc.ac.uk. Accessed on 8 August 2016

Chung MY (2009) Lack of allozyme diversity in populations of the rare, endangered terrestrial orchids *Tipularia japonica* and *Epipactis papillosa* in Korea 278: 203. doi:10.1007/s00606-008-0140-x

Chung MY, Chung MG (2007) Extremely low levels of genetic diversity in the terrestrial orchid *Epipactis thunbergii* (Orchidaceae) in South Korea: implications for conservation. Bot J Linn Soc 155:161-169

Chung MY, Nason JD, Chung MG (2004) Spatial genetic structure in populations of the terrestrial orchid *Cephalanthera longibracteata* (Orchidaceae). Am J Bot 91:52-57

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657-1659

ČNFD (2016) České národní fytocenologické database. Ústavem botaniky a zoologie Přírodovědecké fakulty Masarykovy university. http://quick.florabase.cz. Accessed on 8 August 2016

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. Mol Ecol Resour 10:556-557

CRSF, ZDSF (2012) Info Flora: Das nationale Daten-und Informationszentrum der Schweizer Flora. Chambésy. http://www.crsf.ch. Accessed on 8 August 2016 (in German)

Davies P, Davies J, Huxley A (1983) Wild orchids of Britain and Europe. Chatto and Windus, The Hogarth Press, London

DGRNE (2016) L'Observatoire de la Faune, de la Flore et des Habitats en Wallonie. http://biodiversite.wallonie.be. Accessed on 8 August 2016 (in French)

Didukh YP (2009) Червона книга України. Рослинний. The red Book of Ukraine. Flora. National Sciences of Ukraine, Kiev

Diez JM (2007) Hierarchical patterns of symbiotic orchid germination linked to adult proximity and environmental gradients. J Ecol 95:159-170

Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. Mol Ecol 11:2571-2581

Durka W (1999) Genetic diversity in peripheral and subcentral populations of *Corrigiola litoralis* L.

(Illecebraceae). Heredity 83:476-484

Dusak F, Prat D (2010) Atlas des orchidees de France. Editions Biotope, collection Parthénope, Publications scientifiques du Muséum

Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359-361

Eckert CG, KE Samis, Lougheed SG (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Mol Ecol 17: 1170-1188

Eckstein RL, O'Neill R, Danihelka J, Otte A, Köhler W (2006) Genetic structure among and within peripheral and central populations of three endangered floodplain violets. Mol Ecol 15:2367-2379

Efimov P (2004) Genus *Epipactis* ZINN (Orchidaceae) in the Russia. Turczaninowia 7:8-42 (in Russian)

Ehlers BK, Pedersen H (2000) Genetic variation in three species of Epipactis (Orchidaceae): geographic scale and evolutionary inferences. Biol J Linn Soc 69:411-430

Euro+Med (2011) Euro+Med Plantbase The information resource for Euro-Mediterranean plant diversity. Botanic Garden and Botanical Museum, Berlin. http://ww2.bgbm.org/EuroPlusMed. Accessed on 8 August 2016

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611-2620

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564-567

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479-491

Falush D, Stephens M, Pritchard JK (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics 164:1567-1587

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.

Garza JC, Williamson EG. 2001. Detection of reduction in population size using data from microsatellite loci. Mol Ecol 10:305-318

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995)

Govindaraju DR (1988) Relationship between dispersal ability and levels of gene flow in plants. Oikos 52:31-35

Gustafsson S (2000) Patterns of genetic variation in Gymnadenia conopsea, the fragrant orchid. Mol Ecol 9:1863-1872

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acid S 41:95-98

Hardy OJ, Charbonnel N, Fréville H, Heuertz M (2003) Microsatellite Allele Sizes: A Simple Test to Assess Their Significance on Genetic Differentiation. Genetics 163:1467-1482

Hardy OJ, Vekemans X (2002) spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618-620

Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum Biol 66:591-600

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biol J Linn Soc 58:247-276

Hewitt GM (1999) Post-glacial re-colonization of European biota. Biol J Linn Soc 68:87-112

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. Nature 405:907-913

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philos T Roy Soc B 359:183-195.

Hollingsworth PM, Dickson JH (1997) Genetic variation in rural and urban populations of *Epipactis helleborine* (L) Crantz. (Orchidaceae) in Britain. Bot J Linn Soc 123:321-331

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53:1898-1914

Ivanter EV, Kuznetsov OL (eds) (2007) Red Data Book of the Republic of Karelia. Petrozavodsk, Karelija, pp 368 (in Russian)

Jaarola M, Tegelström H (1995) Colonization history of north European field voles (Microtus agrestis) revealed by mitochondrial DNA. Mol Ecol 4:299-310

Jäkälaniemi A, Crone EE, Närhi P, Tuomi J (2011) Orchids do not pay costs at emergence for prolonged dormancy. Ecology 92:1538-1543

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806

Jakubska-Busse A, Kadej M (2011) The pollination of *Epipactis* Zinn, 1757 (Orchidaceae) species in Central Europe–the significance of chemical attractants, floral morphology and concomitant insects. Acta Soc Bot Pol 80:49-57

Jersáková J, Malinová T (2007) Spatial aspects of seed dispersal and seedling recruitment in orchids. New Phytol 176:237-241

Johnson LA, Soltis DE (1994) matK DNA Sequences and Phylogenetic Reconstruction in Saxifragaceae s. str. Syst Bot 19:143-156

Jost L (2008) GST and its relatives do not measure differentiation. Mol Ecol 17:4015-4026

Julve Ph (ed) (2016) Listes départementales des plantes de France. Version 2015.08 du 30 août 2015. Programme chorologie départementale de tela-botanica. http://www.tela-botanica.org. Accessed on 8 August 2016

Kallio H (2006) The evolution of the Baltic Sea - changing shorelines and unique coasts. Geol S Finl 41:17-21

Kålås JA, Viken Å, Henriksen S, Skjelseth S (eds) (2010) The 2010 Norwegian Red List for Species. Norwegian Biodiversity Information Centre, Norway

Knutsen H, Jorde PE, Gonzalez EB, Robalo J, Albretsen J, Almada V (2013) Climate Change and Genetic Structure of Leading Edge and Rear End Populations in a Northwards Shifting Marine Fish Species, the Corkwing Wrasse (*Symphodus melops*). PLoS ONE 8(6): e67492. doi:10.1371/journal.pone.0067492

Konechnaya GY, Suslova TA (eds) (2004) Red Data Book of the Vologda region. Vol. 2. Plants and Fungi. VGPU Izd. Rus', Vologgda, pp. 360 (in Russian)

Konstantinova NA et al. (eds) (2014) Red Data Book of the Murmansk region. 2nd ed. Azia-print Publishing, Kemerovo, pp. 578 (in Russian)

Kuusk V, Tabaka L, Jankeviciene R (2003) Flora of the Baltic Countries. http://elurikkus.ut.ee. Accessed 8 August 2016

Lammi A P Siikamäki, Mustajärvi K (1999) Genetic diversity, population size, and fitness in central and peripheral populations of a rare plant, Lychnis viscaria. Conserv Biol 13:1069-1078

Lampinen R, Lahti T (2016) Kasviatlas 2015. Helsingin Yliopisto, Luonnontieteellinen keskusmuseo, Helsinki. http://www.luomus.fi/kasviatlas. Accessed on 8 August 2016

Lawton JH (1993) Range, population abundances and conservation. Trends Ecol Evol 8: 409-413

Lesica P, Allendorf, FW (1995) When Are Peripheral Populations Valuable for Conservation?. Conservation Biology 9: 753-760

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452

Lilleleht V (2008) Red data book of Estonia. Threatened Fungi, Plants and Animals. Commission for Nature Conservation of the Estonian Academy of Sciences. http://elurikkus.ut.ee/prmt.php?lang=eng. Accessed on 08 June 2016

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial Genetic Structure of a Tropical Understory Shrub, Psychotria officinalis (Rubiaceae). Am J Bot 82:1420-1425

Machon N, Bardin P, Mazer SJ, Moret J, Godelle B, Austerlitz F (2003) Relationship between genetic structure and seed and pollen dispersal in the endangered orchid *Spiranthes spiralis*. New Phytol 157:677-687

Mantel N (1967) The Detection of Disease Clustering and a Generalized Regression Approach. Cancer Res 27:209-220

Meusel H, Jäger E, Weinert E (1992) Comparative chorology of the Central European flora. Vol. 1, Gustav Fischer Verlag, Jena, Germany, pp 450

Minasiewicz J, Znaniecka J (2014) Characterization of 15 novel microsatellite loci for *Cypripedium calceolus* (Orchidaceae) using MiSeq sequencing. Conserv Genet Resour 6:527-529

Murren C, Ellison A (1998) Seed dispersal characteristics of *Brassavola nodosa* (Orchidaceae). Am J Bot 85:675-680

Natuurpunt (2016) Waarnemingen een initiatief van Natuurpunt Studie vzw en de Stichting Natuurinformatie. www.waarnemingen.be. Accessed on 8 August 2016 (in Dutch)

NBIC, GBIF-Norway (2016) Species Map Service. http://artskart.artsdatabanken.no. Accessed on 8 August 2016

NDFF, FLORON (2015) Verspreidingsatlas. http://www.verspreidingsatlas.nl. Accessed on 8 August 2016 (in Dutch)

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York

NetPhyt (2013) Verbreitungsatlas der Farn- und Blütenpflanzen Deutschlands. Landwirtschaftsverlag, Münster. URL: https://deutschlandflora.de. Accessed on 8 August 2016 (in German)

Nimis PL, Bagella S, Caria MC, Filigheddu R, Moro A, Pittao E, Martello S (2013) An interactive guide to the flora of the temporary ponds of Sardinia (Italy) http://dryades.units.it. Accessed on 8 August 2016

Pandey M., Richards M, Sharma J (2015) Microsatellite-based genetic diversity patterns in disjunct populations of a rare orchid. Genetica 143:693–704

Paquette S R (2012) PopGenKit: Useful functions for (batch) file conversion and data resampling in microsatellite datasets. R package version, 1

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini, D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14:689-701

Petit RJ, Kremer, A, Wagner DB (1993) Finite island model for organelle and nuclear genes in plants. Heredity 71:630-641

Pfeifer M, Schatz B, Picó FX, Passalacqua NG, Fay MF, Carey PD, Jeltsch F (2009) Phylogeography and genetic structure of the orchid *Himantoglossum hircinum* (L.) Spreng. across its European centralmarginal gradient. J Biogeogr 36:2353-2365

Phitos D, Constantinidis T, Kamari G (eds) (2009) The Red Data Book of Rare and Threatened Plants of Greece, Vol. II (E-Z), Hellenic Botanical Society, Patra (in Greek)

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502-503

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions.

Nat Rev Genet 6:847-846

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945-959

Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. Trends Ecol Evol 23:564-571

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. Mol Biol Evol 19:2092-2100

Rankou H (2011) Epipactis atrorubens. The IUCN Red List of Threatened Species 2011: e.T176003A7168400. http://www.iucnredlist.org/details/176003/1. Accessed on 8 August 2016

Rasmussen HN (1995) Terrestrial orchids: from seed to mycotrophic plant. Cambridge University Press

Rasomavièius V, Augustauskas J, Ivinskis P, Motiejûnaitë J, Paltanavièius S & Raudonikis L (eds) (2007) Red data book of Lithuania. (Lietuvos Raudonoji Knyga). Leidykla Lututë. Ministry of Environment of the Republic of Lithuania, Vilnius (in Lithuanian)

Rassi P, Hyvärinen E, Juslén A, Mannerkoski I (2010) The 2010 Red List of Finnish Species. Ympäristöministeriö & Suomen Ympäristökeskus, Helsinki (in Finnish)

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. J Hered 86:248-249

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552-569

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137-138

Rousset F (2008) GENEPOP?007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 8:103-106

Sagarin RD, SD Gaines (2002) The ''abundant centre'' distribution: to what extent is it a biogeographical rule? Ecol Lett 5: 137-147.

SFO (2016) Carte de présence : Carte espèce *Epipactis atrorubens*. http://www.orchisauvage.fr. Accessed on 17 November 2016 (in French)

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American Journal of Botany 94:275-288

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462

Snowarski M (2016) Vascular Plants of Poland photoflora. www.atlas-roslin.pl. Accessed on 8 August 2016 (in Polish)

Soliva M, Widmer A (2003) Gene flow across species boundaries in sympatric, sexually deceptive *Ophrys* (Orchidaceae) species. Evolution 57:2252-2261

Stenström A, Jonsson BO, Jónsdóttir IS, Fagerström T, Augner M (2001) Genetic variation and clonal diversity in four clonal sedges (*Carex*) along the arctic coast of Eurasia. Mol Ecol 10: 497-513

Steward JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: individualistic responses of species in space and time. P Roy Soc B-Biol Sci 277:661-671

Svendsen JI, Astakhov VI, Bolshiyanov DY, Demidov I, Dowdeswell JA, Gataullin V, Hjort C, Hubberten HW, Larsen E, Mangerud J, Melles M, Möller P, Saarnisto M, Siegert MJ (1999) Maximum extent of the Eurasian ice sheets in the Barents and Kara Sea region during the Weichselian. Boreas 28:234-242

Swarts ND, Dixon KW (2009) Perspectives on orchid conservation in botanic gardens. Trends Plant Sci 14:590-598

Taberlet P, Fumagalli L, Wust-Sausy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonisation routes in Europe. Mol Ecol 7:453-464

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17:1105-1109

Taberlet P, Swenson JE, Sandegren F, Bjärvall A (1995) Localization of a contact zone between two highly divergent mitochondrial DNA lineages of the brown bear (*Ursus arctos*) in Scandinavia. Conserv Biol 9:1255-1261

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics

123:585-595

Tałałaj I, Brzosko E (2008) Selfing potential in *Epipactis palustris*, *E. helleborine* and *E. atrorubens* (Orchidaceae). Plant Syst Evol 276:21-29

Templeton AR, Levin DA (1979) Evolutionary Consequences of Seed Pools. Am Nat 114: 232-249

Tikkanen M, Oksanen J (2002) Late Weichselian and Holocene shore displacement history of the Baltic Sea in Finland. Fennia 180:9-20

Tranchida-Lombardo V, Cafasso D, Cristaudo A, Cozzolino S (2011) Phylogeographic patterns, genetic affinities and morphological differentiation between *Epipactis helleborine* and related lineages in a Mediterranean glacial refugium. Ann Bot-London 107:427-436

Tranchida-Lombardo V, Hopkins SE, Selosse M-, Cozzolina S, Taylor DL (2008) Isolation and characterization of new polymorphic microsatellite loci in the mixotrophic orchid *Limodorum abortivum* L. Swartz (Orchidaceae). Mol Ecol Resour 8:1117-1120

Tuulik T (1998) Orchids of Hiiumaa. Pirrujaak 5, Kärdla

Tzvelev NN (ed) (2000) Red Data Book of nature of the Leningrad region. Vol. 2. Plants and Fungi. Mir i Sem'ia, Saint-Petersburg, pp. 672

Vanden Broeck A, Van Landuyt W, Cox K, De Bruyn L, Gyselings R, Oostermeijer G, Valentin B, Bozic G, Dolinar B, Illyés Z, Mergeay J (2014) High levels of effective long-distance dispersal may blur ecotypic divergence in a rare terrestrial orchid. BMC Ecology 14:20 DOI: 10.1186/1472-6785-14-20

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535-538

Westling A (Ed) (2015) Rödlistade arter i Sverige. ArtDatabanken i samarbete med Naturvårdsverket. (in Swedish)

Figure captions

Fig. 1 The sampled regions and the distribution of *Epipactis atrorubens* in Europe. The distribution map is assembled of local distribution maps if available (Kuusk et al. 2003; Efimov 2004; Rasomavièius et al. 2007; BSBI and BRC 2008; Didukh 2009; Phitos et al. 2009; Anthos 2010; Dusak and Prat 2010; CRSF and ZDSF 2012; NetPhyt 2013; Nimis et al. 2013; Lampinen and Lahti 2015; NDFF and FLORON 2015; Artdatabanken 2016; BfN 2016; ČNFD 2016; DGRNE 2016; Julve 2016; Natuurpunt 2016; NBIC and GBIF-Norway 2016; SFO 2016; Snowarski 2016) or from European distribution maps (Euro+Med 2011; Rankou 2011). The sampled populations are marked with a black square.

Fig. 2 (a) The distribution and frequencies of the combined chloroplast haplotypes in the populations. (b) Maximum parsimony network of the combined chloroplast loci. The sizes of the haplotype rings reflects the haplotype frequency among all individual samples. Each branch indicates one mutation step and colours indicate the distribution of the haplotype among the region. The number of samples (N) with the corresponding haplotype are shown below the haplotype name.

Fig. 3 (a) The likelihood values (Ln P(D) in black) with their standard deviations and the ΔK values (in grey) for the different number of clusters (K) and the assignment probabilities for each sampled individual for (b) K=2 and (c) K=4 given as barplots on the map.

Table 1 Diversity estimates with their standard deviations for the combined chloroplast sequence loci. Number of samples with successful amplification (N), number of haplotypes (H), number of segregation sites (S), haplotype diversity (Hd), nucleotide diversity (π) and Watterson's theta (Θ_S). No chloroplast genetic variation was found in Hanko and Juankoski and therefore the diversity values are not given.

	N	H	S	Hd ± S.D.	π ± S.D.	Θ_S ± S.D.
Oulanka	72	7	4	0.621±0.056	0.0013±0.0009	0.0008±0.0005
Juankoski	25	1	-	-	-	-
Hanko	19	1	-	-	-	-
Estonia	52	3	3	0.597±0.042	0.0014±0.0010	0.0007±0.0004
Karelia	24	2	1	0.507±0.045	0.0005±0.0005	0.0003±0.0003
Pskov	19	3	3	0.292±0.127	0.0007±0.0006	0.0009±0.0005
Vologda	15	3	3	0.686±0.068	0.0015±0.0011	0.0009±0.0006

Table 2 Genetic diversity estimates with their standard deviations for the nuclear loci. The sample size (N), expected heterozygosity (H_E), observed heterozygosity (H_O), allelic richness (A_R), inbreeding coefficient (F_{IS}) and the p-value for Hardy-Weinberg equilibrium P_{HW}, the p-value for the Wilcoxon sign rank test (P_W) and the Garza Williamson Index (G-W). The indicative adjusted nominal level (0.05) is 0.0003. H_E, H_O, A_R and their standard deviations or confidence intervals are calculated as means over loci. H_E, H_O, A_R are calculated using the resampling method.

Population	N	H_O (C.I.)	H_E (C.I.)	$A_R \pm SD$	F_{IS}	P_{HW}	P_W	G-W
Oulanka	100	0.225 (0.204-0.290)	0.238 (0.308-0.387)	2.324±0.588	0.307	**0.001**	0.016	0.437±0.175
Hanko	24	0.233 (0.229-0.356)	0.281 (0.195-0.330)	1.899±0.498	0.005	0.466	0.016	0.436±0.140
Juankoski	30	0.292 (0.238-0.348)	0.314 (0.217-0.382)	2.029±0.191	0.197	**0.045**	0.016	0.483±0.119
Estonia	60	0.397 (0.364-0.463)	0.371 (0.339-0.407)	2.123.±0.371	-0.080	0.080	0.016	0.451±0.197
Karelia	29	0.293 (0.263-0.364)	0.259 (0.236-0.301)	1.853±0.873	-0.133	0.083	0.016	0.357±0.193
Pskov	30	0.128 (0.071-0.167)	0.123 (0.070-0.150)	1.445±0.490	-0.027	0.574	0.016	0.397±0.287
Vologda	17	0.268 (0.201-0.311)	0.206 (0.156-0.220)	1.548±0.629	-0.289	**0.021**	0.016	0.556±0.139

Table 3 The pairwise Φ_{ST}-estimates among regions for the combined chloroplast loci (below diagonal) and the F_{ST} estimates for the nuclear microsatellite loci (above diagonal). Significant (p≤0.05) values are shown in bold.

	Oulanka	Juankoski	Hanko	Estonia	Karelia	Pskov	Vologda
Oulanka	-	0.022	-0.010	**0.074**	-0.032	**0.068**	0.026
Juankoski	**0.134**	-	-0.089	-0.149	-0.011	**0.253**	**0.176**
Hanko	**0.120**	0.000	-	**0.059**	**0.119**	**0.235**	0.036
Estonia	**0.075**	**0.336**	**0.314**	-	**0.167**	**0.227**	**0.155**
Karelia	**0.245**	**0.571**	**0.535**	**0.418**	-	**0.177**	-0.025
Pskov	0.008	0.104	0.079	**0.159**	**0.382**	-	**0.077**
Vologda	**0.192**	**0.633**	**0.590**	0.065	**0.591**	**0.340**	-