Tree Genetics & Genomes (2018) 14: 19
https://doi.org/10.1007/s11295-018-1235-3

ORIGINAL ARTICLE

High gene flow and complex treeline dynamics of *Larix* Mill. stands on the Taymyr Peninsula (north-central Siberia) revealed by nuclear microsatellites

S. Kruse[1,2] · L. S. Epp[1] · M. Wieczorek[1,3] · L. A. Pestryakova[4] · K. R. Stoof-Leichsenring[1] · U. Herzschuh[1,2,3]

Received: 23 June 2017 / Revised: 29 November 2017 / Accepted: 24 January 2018 / Published online: 7 February 2018

Abstract

Arctic treelines are facing a strong temperature increase as a result of recent global warming, causing possible changes in forest extent, which will alter vegetation-climate feedbacks. However, the mode and strength of the response is rather unclear, as potential changes are happening in areas that are very remote and difficult to access, and empirical data are still largely lacking. Here, we assessed the current population structure and genetic differentiation of *Larix* Mill. tree stands within the northernmost latitudinal treeline reaching ~ 72° N in the southern lowlands of the Taymyr Peninsula (~ 100° E). We sampled 743 individuals belonging to different height classes (seedlings, saplings, trees) at 11 locations along a gradient from 'single tree' tundra over 'forest line' to 'dense forest' stands and conducted investigations applying eight highly polymorphic nuclear microsatellites. Results suggest a high diversity within sub-populations (H_E = 0.826–0.893), coupled, however, with heterozygote deficits in all sub-populations, but pronounced in 'forest line' stands. Overall, genetic differentiation of sub-populations is low (F_{ST} = 0.005), indicating a region-wide high gene flow, although 'forest line' stands harbour few rare and private alleles, likely indicating greater local reproduction. 'Single tree' stands, located beyond the northern forest line, are currently not involved in treeline expansion, but show signs of a long-term refuge, namely asexual reproduction and change of growth-form from erect to creeping growth, possibly having persisted for thousands of years. The lack of differentiation between the sub-populations points to a sufficiently high dispersal potential, and thus a rapid northward migration of the Siberian arctic treeline under recent global warming seems potentially unconstrained, but observations show it to be unexpectedly slow.

Keywords Larch · Population genetics · Boreal forests · Tundra-taiga transition · Range expansion

Communicated by Y. Tsumura

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-018-1235-3) contains supplementary material, which is available to authorized users.

✉ S. Kruse
stefan.kruse@awi.de

[1] Polar Terrestrial Environmental Systems Research Group, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Telegrafenberg A43, 14473 Potsdam, Germany

[2] Institute of Biology and Biochemistry, University of Potsdam, 14476 Potsdam, Germany

[3] Institute of Earth and Environmental Science, University of Potsdam, 14476 Potsdam, Germany

[4] Institute of Natural Sciences, North-Eastern Federal University of Yakutsk, 677000 Yakutsk, Russia

Introduction

Temperatures in the Arctic strongly increased during recent decades (IPCC 2013), which is unanimously expected to cause a northward transition of the latitudinal treeline (Holtmeier and Broll 2007; MacDonald et al. 2008). The timing, spatial pattern, and regional peculiarities of such forest expansion are, however, still under debate (Holtmeier and Broll 2005; Harsch et al. 2009). Such knowledge is of particular interest because the proposed vegetation turnover of tundra into forest areas will decrease the albedo causing a positive feedback to regional warming which could amplify the functioning of emerging forests as carbon sinks (Bonan 2008).

The Siberian treeline is formed by largely mono-specific forest canopies (Abaimov 2010) of *Larix*, i.e. *Larix sibirica* (60–90° E), *L. gmelinii*, (90–120° E), and *L. cajanderi* (120–175° E), going from west to east. The *Larix* forests probably

expanded during the late glacial and the first half of the Holocene (MacDonald et al. 2000a) from many small glacial refugia (Polezhaeva et al. 2010, 2013; Semerikov et al. 2013), a pattern that was also recently observed for *Picea* in western Eurasia (Tsuda et al. 2016). Palaeobotanical evidence indicates that the treeline on the Taymyr Peninsula was located about 250 km to the north of its present position during the mid-Holocene (Bigelow et al. 2003). In response to late-Holocene cooling, the treeline retreated until reaching an overall minimum about 2000 years ago (MacDonald et al. 2000a, 2008; Klemm et al. 2016). Vegetation plot analyses on the southern Taymyr Peninsula revealed that tree stands at the forest line have recently densified, probably in response to warming starting about AD 1900 (Wieczorek et al. 2017). In contrast, recruitment is largely absent in the northernmost 'single tree' stands, which are mostly krummholz growth-form, and are dispersed up to several tens of kilometres ahead of the current forest line. It is currently unknown whether these krummholz trees represent recent colonisation events, which would indicate a high speed of migration, or, conversely, they represent relict forest stands of past warm phases (Corlett and Westcott 2013).

The investigation of the genetic structure of populations has the potential to unveil the spatial pattern of forest expansion (Arenas et al. 2012), but to date, only a few case studies have made use of this tool to analyse currently expanding forests (e.g. Piotti et al. 2009; Pluess 2011). If the foremost tree stands originated from newly founded populations, they should bear genetic similarities with those expected at the moving front of the forest line. In such a situation, recruitment would mostly be from a few individuals, thus probably causing an inbred population (Davis et al. 2015). Through founder effects, marginal populations, initiated from a limited number of reproductive individuals, should be characterised by a decrease in genetic diversity and differentiation between different populations (Hartl and Clark 2007). This has been demonstrated in a range-wide genetic structure analysis of *Picea jezoensis* with nuclear microsatellite markers (Aizawa et al. 2009). Alternatively, if the foremost tree stands represent relict stands of an earlier continuous distribution range, the genetic pattern depends on the mode of reproduction. If sexual reproduction among the relatively small sub-population of the relict tree stands is predominant, genetic diversity should have diminished and differences among separate 'single tree' stand sub-populations should have accumulated over time (Hartl and Clark 2007), but if reproduction is clonal, the past high diversity state of the population should have been conserved. Empirical evidence from treeline habitats for such processes is currently lacking. Generally, a reduction in genetic diversity and genetic differentiation of populations is counterbalanced by processes that retain gene flow such as pollen dispersal (Burczyk et al. 2004; Nishimura and Setoguchi 2011). The extreme case of this 'rescue effect' is panmictic gene flow, i.e. genetic exchange over broad areas is higher than processes that lead to a fixation of differentiation (namely genetic drift and inbreeding) so that no differentiation can accumulate at the most distant sub-populations (Ray and Excoffier 2010).

Microsatellites are well suited for small-scale population studies as they allow for the identification of high-resolution relatedness patterns (Selkoe and Toonen 2006; Ashley 2010). Several microsatellite primers are available for the investigation of *Larix decidua* (Wagner et al. 2012) and *Larix kaempferi* populations (Isoda and Watanabe 2006). The latter primers were recently applied in population studies of closely related larch species. Using these markers, Oreshkova et al. (2015), for example, find highly inbred populations of *L. cajanderi* on the Kamchatka Peninsula, which are furthermore characterised by high genetic differentiation from stands located on mainland Asia. High-resolution population studies of *Larix* in Asia have, to date, only been performed in the Russian Far East (Oreshkova et al. 2015) and the southern boreal areas (Nishimura and Setoguchi 2011), while from northern boreal areas, only studies that focus on large-scale post-glacial dynamics of *Larix* and speciation using mostly cytoplasmic markers or allozyme loci are available (Semerikov et al. 1999, 2013; Semerikov and Lascoux 2003; Semerikov and Polezhaeva 2007; Polezhaeva et al. 2010, 2013, Oreshkova et al. 2013, 2014, 2015). Recent genetic studies of boreal forest taxa revealed the past population demography using nuclear microsatellite marker (*Picea* and *Betula*, Tsuda et al. 2016; Tsuda et al. 2017). However, a fine-scale reconstruction of the migration pattern at the latitudinal treeline is currently lacking, but is possible when using microsatellite markers.

Here, we aim to reconstruct the population dynamics on the southern Taymyr Peninsula (north-central Siberia, Russia)—an area that is characterised by particularly strong warming rates (IPCC 2013). We applied microsatellites to detect population dynamics within a single region and pursue the following questions: *does genetic diversity within and genetic distance among sub-populations located across the treeline ecotone reveal a distinctive structure or rather panmictic gene flow? If a structure can be observed, is the northernmost forest characterised by a northward moving forest-front established by recent recruitment from long-distance dispersal, or by the persistence of remnants of a formerly widespread forest, or a combination of both?*

Materials and methods

Study area

The study area comprises the lowlands on the southern Taymyr Peninsula and in the north of the Putorana Plateau. During the last glacial maximum, the area was not overlain by

glaciers but was covered by arctic tundra (Svendsen 2004; Wetterich et al. 2011). The area is underlain by continuous permafrost and characterised by organic-rich cryosols with a shallow active-layer depth of up to 0.6 m (Park et al. 2016). Its topography is flat with gentle hills that seldom reach more than 200 m a.s.l. (Fig. 1). Thermokarst lakes and polygonal wetlands have developed in the depressions, particularly along the rivers.

Climate is continental sub-arctic. Temperatures reach a maximum of ~ 12.5 °C in July, which is the only frost-free month throughout the year (1960–1990, Khatanga weather station at 71.983° N, 102.466° E). Coldest temperatures are measured in January, regularly falling below − 30 °C. Annual precipitation sums up to 270 mm, of which 38% falls between June and August.

The vegetation is characterised by an extensive moss and lichen cover (5–10 cm thick) mainly *Sphagnum*, *Hylocomium*, and *Aulacomnium*. The herbaceous and dwarf-shrub layer of up to around 50 cm height is formed by sedges (e.g. *Eriophorum, Carex*) and shrubs (*Betula nana*, *Salix pulchra*, *Ledum palustre*, *Vaccinium* spp., *Alnus viridis* ssp. *fruticosa*) (Zyryanova et al. 2010; Zibulski et al. 2016). In sheltered places, open forest-tundra stands grow at varying densities which turn into open woodland and northern taiga forests towards the south (Kharuk et al. 2006, 2013; Zyryanova et al. 2010). *Larix* can endure unfavourable conditions as a

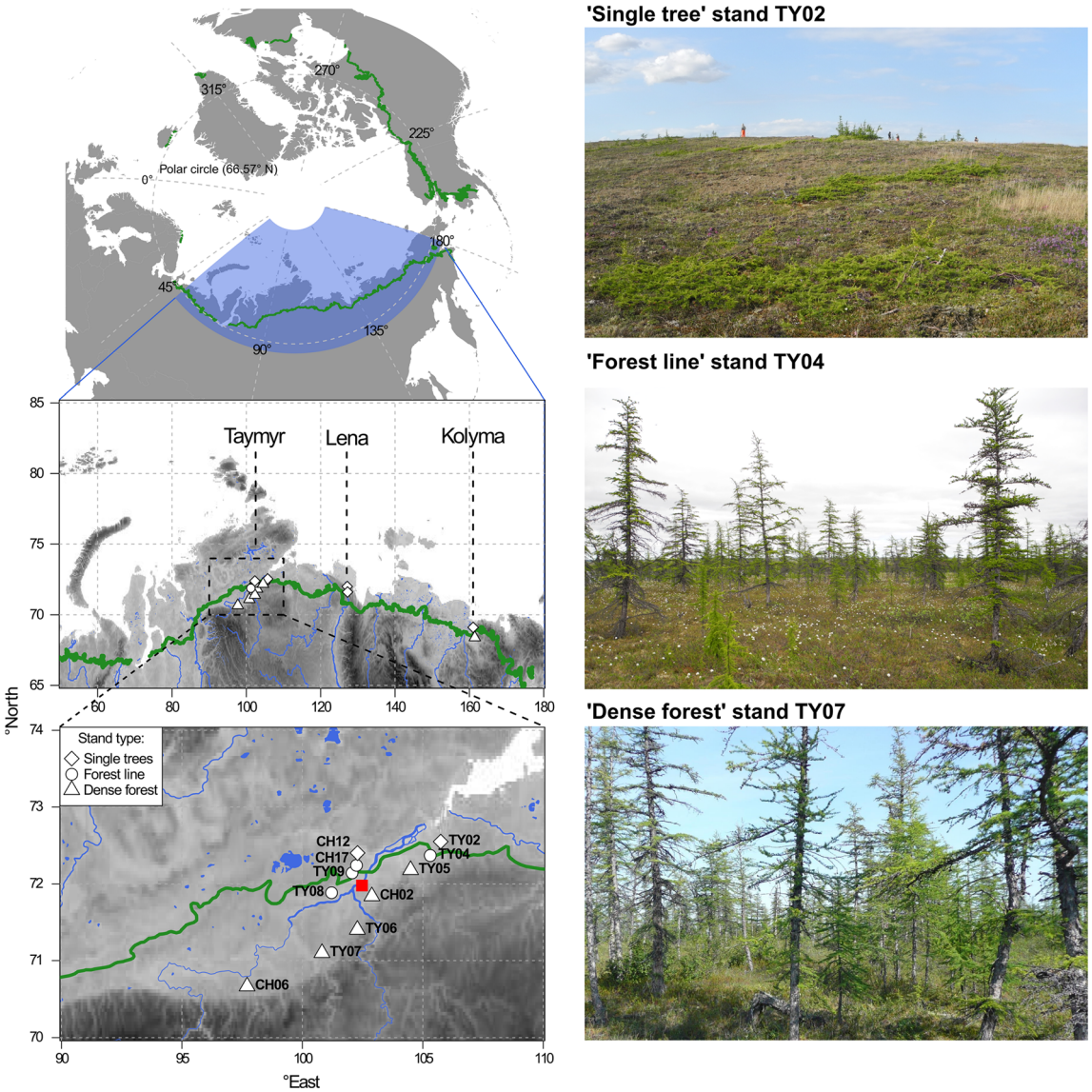

Fig. 1 Overview of the study area located in the vicinity of the city of Khatanga (red square) on the Taymyr Peninsula. Pictures are given for typical larch stands of the three distinguished categories. The green line marks the modern maximum position of the treeline (Walker et al. 2005). Elevation ranges between 1 and 2521 m (WorldClim1.4, Hijmans et al. 2005). Rivers and lakes are given in blue colours (GSSHS updated version 2.2.2 01.01.2s013 first published by Wessel and Smith 1996)

creeping growth-form ('krummholz') at favourable microsites in the tundra (own observation and MacDonald et al. 2008).

Sampling

Samples of *Larix gmelinii* (Rupr.) Rupr. were collected in the summers of 2011 and 2013 from 11 sites located in the forest-tundra transition area on the southern Taymyr Peninsula in north-central Siberia (~ 70° N, ~ 100° E). Sampling sites were remote with respect to riverine and human impact and were reached by helicopter. The sites span an area that extends approximately 220 km from south to north (70.7° to 72.4° N) and 300 km from west to east (97.7° to 105.4° E). They were selected to represent even vegetation stands. Differences in the number of sampled individuals were due to timetable restrictions, i.e. some sites close to the campsites, which were set up for multiple days, were intensively sampled, while others were reached during short helicopter stops (Table 1, further details in the Electronic supplementary information S2.2). At intensively sampled sites, trees were sampled within a c. 100 × 100 m square, focusing mainly on the central 20 × 20 m within a homogeneous forest where other vegetation analyses were also performed (Wieczorek et al. 2017). This sampling scheme was adapted in the low tree densities at the krummholz sites, namely CH12 and TY02, where larger areas were sampled (all individuals sampled within c. 150 × 150 m and 80 × 50 m). At each site, we sampled from three different height classes selecting a minimum of ten individuals each where available (Table 1) to evaluate responses between current and past recruitments. For evaluating the strength of differentiation on a local to regional scale, ten *L. gmelinii* individuals were collected from the Lower Lena area (~ 70° N, ~ 120° E) in 2014 and ten *L. cajanderi* individuals were collected from two locations in the Kolyma river basin (~ 60° N, ~ 161° E). These species are genetically close and were only recently differentiated (e.g. Abaimov 2010). All needle samples were dried and stored on silica gel in the field and stored at 8 °C in the Alfred Wegener Institute in Potsdam, Germany until processing. In total, 763 individuals were sampled for which the identifier, geographic position and height were recorded. Sampling locations, morphological data and microsatellite genotype data are available at https://doi.pangaea.de/10.1594/PANGAEA.870947.

DNA extraction

Approximately 100 mg of dried needles were shock frozen in liquid nitrogen and ground with a FastPrep®-24 instrument (MP BIOMEDICALS). DNA extractions were carried out using the Invisorb® Spin Plant Mini Kit (STRATEC MOLECULAR) or DNeasy Plant Mini Kit (QIAGEN), following the supplier's protocol, with modifications.

Selection of microsatellite primers and design of multiplexes

After literature searches (*Larix lyallii* and *L. occidentalis*, Khasa et al. 2000; *L. kaempferi*, Isoda and Watanabe 2006; *L. occidentalis*, Chen et al. 2009; *L. decidua*, Wagner et al. 2012), 34 primers designed for *Larix* (Table 2) were tested in single-primer pair PCR on a test set of eight individuals at four locations. These primers were reported to show cross-species amplification. Specificity for the microsatellite region was checked for the most promising primers by cloning and sequencing the resulting PCR products (results not shown).

Of the primers tested, 16 primer pairs were optimised for the full analysis, following multiplex reactions. Optimal combinations of primer pairs, with their expected sizes (Table 2) as published, were made with the aid of Multiplex Manager (Holleley and Geerts 2009). Four different fluorescent dyes—6-FAM™, VIC®, NED™ and PET™ (LIFE TECHNOLOGIES)—were used to label primers and amplified using a nested (multiplex) PCR approach with Q-tails to reduce costs (Schuelke 2000; Micheneau et al. 2011). The three final multiplexes were designed after empirically testing duplexes for amplification strength and specificity of the reaction (Table 2).

PCR reactions and fragment length estimation

PCR reactions of 10 μl were set up using the Multiplex PCR Master Kit (QIAGEN), containing ~ 1–5 ng template DNA, primer (0.7 μM forward, 1.0 μM reverse and 1.0 μM fluorescent-labelled reverse primer for Q1-Q4 tags; Table 2) and MasterMix (containing 25 units HotStarTaq® DNA polymerase, Multiplex PCR Buffer (pH 8.7) with a final 3 mM $MgCl_2$ and dNTP Mix). PCRs were performed using the following reaction profile: 15 min activation of polymerase and initial DNA melting at 94 °C, followed by 20 cycles of 30 s 94 °C, 90 s 60 °C and 15 s 72 °C, a further 10 repeats for 30 s 94 °C, 45 s 53 °C and 15 s 72 °C and a final elongation step at 60 °C for 45 min to minimise stutter of bands introduced by incomplete A-tailing of products (QIAGEN manual for Multiplex PCR Master Kit).

Fragment length estimation was performed by SourceBiosciences (Oxford, UK), with 1 μl of each PCR product added to 9 μl HiDi formamide/size standard mixture LIZ 600®, dye set DS-33 and run on an ABI 3730xl DNA Analyser. The resulting raw data were processed in Geneious (version 7.1.5, Biomatters Ltd.) using the Microsatellite plugin (version 1.4.0, Biomatters Ltd.), and the first automatically scored peaks were checked

Table 1 Sampled sub-populations in northern Siberia

Region	Sub-population name	LAT (°E)	LON (°N)	Basal area (m^2/ha)	Tree stand type	Total number of samples	Samples ≥ 2 m	Samples < 2 and ≥ 0.4 m	Samples < 0.4 m
Taymyr Peninsula (*Larix gmelinii*)	TY02	72.55	105.74	0.02	'Single trees'	49	5	33	11
	CH12	72.40	102.29	0.20	'Single trees'	22	8	12	2
	TY04	72.37	105.31	2.30	'Forest line'	131	78	27	26
	CH17	72.24	102.25	6.70	'Forest line'	62	19	22	21
	TY09	72.14	102.06	8.50	'Forest line'	189	99	22	68
	TY08*	71.89	101.21	–	'Forest line'	30	24	6	–
	TY05*	72.18	104.49	–	'Dense forest'	31	25	5	–
	CH02	71.84	102.88	9.00	'Dense forest'	30	11	9	10
	TY06*	71.40	102.28	–	'Dense forest'	44	38	6	–
	TY07	71.10	100.81	10.80	'Dense forest'	131	67	19	45
	CH06	70.66	97.71	12.00	'Dense forest'	24	14	–	10
Lower Lena (*Larix gmelinii*)	LLL	71.62	127.23	–	'Single trees'	5	5	–	–
	LLR	72.01	127.13	–	'Single trees'	5	5	–	–
Lower Kolyma (*Larix cajanderi*)	KO022	68.39	161.45	–	'Dense forest'	5	5	–	–
	KO05	69.11	161.01	–	'Single trees'	5	5	–	–

Sampling dates of sites beginning with 'CH' were in summer 2011, 'KO' in 2012 and 'TY' and 'LL' in 2013 (Wieczorek et al. 2017)

*Locations with individuals sampled during a short helicopter stop

manually. Subsequent length estimation of scored alleles was calculated in R using a locally weighted method. Bins were constructed on the basis of the observed alleles and a model built for each locus manually to optimise binning (not shown).

Data processing and analyses

Sample preselection

Allelic data were imported into R (version 3.2.2) (R Core Team 2015) using the package 'adegenet' (version 1.4-2) (Jombart 2008; Jombart and Ahmed 2011). Prior to analyses we excluded samples which originated from clonal growth or were closely related to each other. To do this, we calculated the absolute allele differences between all 16 loci ('diss.dist' function) to compare the alleles of each pair of individuals and each locus, and if the sum is 32, individuals are different at all 16 loci. We kept the largest individual of each pair that was only differentiated by five or less alleles. Of the 865 samples, we retained 763 individuals and excluded 102 samples of clones or very closely related individuals (not shown).

We sub-sampled the populations, producing 100 datasets, to minimise the effect of unequal sample sizes on *F* statistics and derivatives. Regarding the three different height classes (seedling < 0.4 m, sapling 0.4–2 m, tree > 2 m), we sampled nine individuals according to the minimum number of nine samples taken at intensively sampled sites, although this was not possible at site CH06 where this size class was completely missing (Table 1). For each set, we calculated descriptive statistics and report their mean values and standard deviations throughout the manuscript unless otherwise stated. Furthermore, we checked the influence of the distance of samples in each sub-population on the number of alleles (Electronic supplementary information S2.2) and found no significant systematic effect within each sub-population, but could not completely rule out any influences that the different distances might have had on the results.

Evaluation of the microsatellite loci

Of the 16 markers analysed, those that departed from the Hardy-Weinberg equilibrium (HWE) in all populations and showed significant linkage disequilibrium ('LD2' function from pegas-library in R; $p < 0.05$) were excluded from the analyses (Table 2; Fig. 2S in the Electronic supplementary information; following Semerikov et al. 1999). Furthermore, high inbreeding coefficient values (F_{IS}) may indicate primer-site mutation or possibly large allele drop-out or slippage when PCR reactions are carried out (e.g. Chapuis and Estoup 2007). The frequency of null alleles is associated with allelic dropout (Wattier et al. 1998; Miller and Waits 2003), mutations in primer binding site (Shaw et al. 1999), and scoring errors due to stutter or low DNA concentrations (Shinde et al. 2003; Wandeler et al. 2003). In consequence, we excluded loci with high frequencies of null alleles, and, with significant inbreeding coefficients, inferred by INEST2 to account for the presence of null alleles (Chybicki and Burczyk 2008).

The frequency of null alleles was calculated as the mean value in all populations from the Taymyr Peninsula directly

Table 2 Summary table for all 16 tested loci for the 11 sub-populations from the Taymyr Peninsula

No.	Locus[a]	Multiplex[b] (expected fragment length (bp))	TAG[c]	Taymyr Peninsula							Taymyr-Kolyma-Lena F_{ST} (region)
				Observed fragment length (bp)	Number of alleles	H_O	H_E	Significant HWE deviations	F_{IS}	F_{ST}	
1	***bcLK253***	1 (219–242)	Q3	211–247	**16.99 ± 0.39**	**0.827**	**0.854**	**1.6 ± 0.9**	**0.006**	**0.010 ± 0.003**	**0.022 ± 0.011**
2	***Ld101***	1 (203–335)	Q4	196–236	**15.74 ± 0.79**	**0.545**	**0.766**	**6.6 ± 1.3***	**0.024**	**0.009 ± 0.004**	**0.072 ± 0.017**
3	***bcLK228***	2 (183–215)	Q4	133–269	**18.70 ± 0.66**	**0.697**	**0.905**	**4.3 ± 1.1***	**0.010**	**0.005 ± 0.003**	**0.077 ± 0.012**
4	***bcLK189***	3 (152–188)	Q2	152–242	**33.39 ± 1.50**	**0.743**	**0.910**	**3.0 ± 1.3***	**0.023**	**0.004 ± 0.002**	**0.009 ± 0.010**
5	***bcLK211***	1 (197–220)	Q2	194–250	**22.97 ± 1.09**	**0.835**	**0.917**	**1.6 ± 1.2***	**0.020**	**0.004 ± 0.002**	**0.032 ± 0.009**
6	***Ld42***	3 (186–196)	Q4	187–201	**7.86 ± 0.35**	**0.520**	**0.783**	**7.1 ± 1.2***	**0.023**	**0.003 ± 0.003**	**0.102 ± 0.035**
7	***bcLK056***	2 (166–212)	Q1	154–256	**31.79 ± 1.05**	**0.648**	**0.940**	**9.4 ± 1.0***	**0.010**	**0.002 ± 0.002**	**0.018 ± 0.009**
8	***bcLK263***	2 (208–264)	Q2	198–280	**39.77 ± 0.96**	**0.868**	**0.953**	**2.4 ± 0.9**	**0.017**	**0.001 ± 0.002**	**0.006 ± 0.006**
9	bcLK224	2 (139–151)	Q3	140–158	8.55 ± 0.58	**0.261**	**0.650**	10.4 ± 0.7*	−0.008	0.011 ± 0.008	−0.002 ± 0.024
10	Ld56	3 (252–269)	Q2	227–277	16.74 ± 0.87	**0.373**	**0.836**	9.8 ± 0.4*	0.011	0.011 ± 0.005	0.023 ± 0.016
11	bcLK225	3 (176–200)	Q1	171–225	23.37 ± 0.93	**0.307**	**0.895**	11.0 ± 0.1*	−0.027	0.006 ± 0.003	0.136 ± 0.019
12	bcLK235	3 (179–225)	Q3	133–269	32.14 ± 1.04	**0.755**	**0.920**	4.7 ± 1.1*	0.012	0.005 ± 0.002	0.021 ± 0.009
13	bcLK260	2 (119–131)	Q1	101–155	14.25 ± 0.80	**0.382**	**0.848**	10.9 ± 0.4*	−0.020	0.002 ± 0.003	0.051 ± 0.019
14	bcLK241	1 (158–196)	Q1	194–250	22.10 ± 1.13	**0.407**	**0.890**	10.2 ± 0.6*	0.010	0.001 ± 0.004	0.041 ± 0.016
15	Ld45	2 (221–237)	Q3	214–260	12.07 ± 0.87	**0.327**	**0.758**	10.6 ± 0.6*	−0.007	−0.004 ± 0.004	0.051 ± 0.020
16	bcLK066	1 (153–168)	Q3	154–182	10.70 ± 0.46	**0.431**	**0.707**	6.2 ± 1.2*	0.068[d]	−0.007 ± 0.003	0.006 ± 0.017

The final retained set of eight loci is in bold and sorted by decreasing population genetic differentiation value F_{ST}. Observed and expected heterozygosity is given by H_O and H_E. F_{IS} values were estimated with INEST2 and account for the presence of null alleles. Weir and Cockerham's F_{ST} for the three regions—Taymyr Peninsula, lower Lena and lower Kolyma—is also given

*$p > 0.001$, significant deviations from the Hardy-Weinberg equilibrium (HWE); see 'Materials and methods' section for details

[a] Locus—marker names beginning 'bcLK' are developed by Isoda and Watanabe (2006) and those with 'Ld' by Wagner et al. (2012)

[b] Multiplex—number indicates the three primer mixes applied in a simultaneous PCR

[c] TAG, tailing sequence at forward primer: Q1, TGTAAAACGACGGCCAGT (Schuelke 2000); Q2, TAGGAGTGCAGCAAGCAT; Q3, CACTGCTT AGAGCGATGC; Q4, CTAGTTATTGCTCAGCGGT (Q2–Q4, after Culley et al. 2008); fluorescent-labelled reverse primers for each tag were Q1, 6-FAM™; Q2, NED™; Q3, VIC®; and Q4, PET™

[d] Values exceeding the 97.5% confidence interval

from the allele data with the 'info_table' function from the poppr-library in R. Further methods to estimate the null allele frequency were carried out in CERVUS (Marshall et al. 2009), ML-NULL (Kalinowski and Taper 2006), GENEPOP (Rousset 2008) and MICRO-CHECKER (Van Oosterhout et al. 2004). The methods showed similar frequencies for three randomised datasets and ranged from close to zero up to 38% (Fig. 1S in the Electronic supplementary information). We used the mean value as the frequency of null alleles for our evaluation of microsatellite loci.

Population genetic structure

For each locus and sub-population we calculated the observed and expected heterozygosity, tested for Hardy-Weinberg equilibrium and made exact tests based on 1000 Monte Carlo permutations of alleles ('hw.test' function from pegas-library in R). Allele set diversity in groups was assessed by allelic richness, rare alleles (frequency < 1% in total dataset) and private alleles. *F* statistics (F_{ST}, F_{IS}, F_{IT}) of Weir and Cockerham (1984) were calculated for groups.

To calculate the average inbreeding coefficient of a sub-population relative to all individuals, we generated pairwise population differentiation measures, as well as general metrics for all sub-populations: Weir and Cockerham's F_{ST} (Weir and Cockerham 1984), Nei's G_{ST} (Nei 1973, 1977), Hedrick's G'_{ST} (Hedrick 2005), and Jost's *D* (Jost 2008). Additionally, the inbreeding measure F_{IS} was calculated from the number of inbred individuals in the local sub-population. The degree of differentiation of sub-populations was further analysed by constrained ordination analysis, namely distance-based redundancy analysis (dbRDA) using the 'capscale' function in the vegan library (Oksanen et al. 2016), on the entire subset of 11 sub-populations on the Taymyr Peninsula. The significance of the dbRDA was tested with a permutation test of 999 iterations ('anova.cca' function in the vegan library). We used Bruvo's genetic distance as a pairwise distance measure between individuals, which calculates the mean of stepwise distances between alleles for each individual locus (Bruvo et al. 2004; 'bruvo.dist' function in the poppr library: Kamvar et al. 2014, 2015). Groups of values were compared using the pairwise Wilcoxon rank sum test with a Bonferroni correction for multiple testing.

Finally, we assigned the group members to the sub-populations and analysed the assignment probability of each individual to these clusters with a discriminant analysis of principal components (DAPC) (Jombart et al. 2010). This was also done for the different height classes (seedling, sapling, tree).

Furthermore, cluster and spatial or temporal inference methods (PCA, DAPC, k-means-clustering, STRUCTURE, Geneland, EEMS, DIYABC) were applied but produced no significant or conclusive pattern in general, due to the very high variation (detailed information on implementation and results can be found in the Electronic supplementary information S5, S6 and S7).

Results

Evaluation of the microsatellite loci for the Taymyr dataset

All investigated 16 loci are highly polymorphic, and the Taymyr dataset, after resampling to create a dataset with equal sizes (see above), has average values across sub-populations of 7.9 to 39.8 different alleles per locus (Table 2). Observed heterozygote frequencies range between 0.261 and 0.868 and are generally lower than the expected heterozygosity of between 0.650 and 0.953. The highest difference between observed and expected heterozygosity was found at locus bcLK225 (0.588). Observed frequencies deviated significantly from the Hardy-Weinberg equilibrium for all but two loci (bcLK253 and bcLK263, exact test based on 1000 permutations; Table 2; Table 2S in the Electronic supplementary information).

Missing alleles total, on average, to 3.1% (Table 1S in the Electronic supplementary information). The highest amounts were found for loci bcLK241 (13%), Ld56 (13%), bcLK225 (9%), bcLK224 (5%) and Ld45 (5%): all others have values < 2% (Fig. 1S in the Electronic supplementary information). The amount of missing alleles is evenly distributed among populations, but for three loci, these are above 20% at a single site (bcLK241 in CH06, Ld56 in CH12 and Ld56 in CH02; Table 1S in the Electronic supplementary information). Significant ($p < 0.05$) linkage was revealed for four pairs of loci (bcLK066 and bcLK260, bcLK066 and bcLK263, bcLK253 and bcLK235, bcLK263 and bcLK225) and for a further three pairs at a significance level of $p < 0.1$ (bcLK253 and bcLK263, Ld101 and bcLK189, bcLK235 and bcLK263) (Fig. 2S in the Electronic supplementary information).

We excluded the loci bcLK241, Ld56, bcLK225, bcLK224, bcLK260 and Ld45 because they have high F_{IS} (all > 0.550; Table 2), as well as many significant deviations from the Hardy-Weinberg equilibrium (Table 2S), and all appeared to have a high null allele frequency (all > 25%, maximum c. 35%; Fig. 1S in the Electronic supplementary information). Furthermore, bcLK235 and bcLK066 were excluded due to their linkages with bcLK263 and relatively high F_{IS}, null allele frequencies and Hardy-Weinberg equilibrium deviations in some populations compared with bcLK263. Finally, we retained eight loci for further analyses (bcLK253, bcLK211, Ld101, bcLK056, bcLK263, bcLK228, bcLK189 and Ld42; Table 2).

Population differentiation

Differentiation among populations throughout northern Siberia

The regional differentiation among the populations from the Taymyr Peninsula, the lower Lena and the lower Kolyma is highly significant ($F_{ST} = 0.042$ for 8 loci, $p < 0.001$; Table 2; Fig. 2). The locus Ld42 contributes most strongly to the separation of the regions with $F_{ST} = 0.102$. The separation between the Kolyma region, in which the forests are composed of *L. cajanderi*, and the other two regions, which are in the *L. gmelinii* range, is strongest. Pairs of Kolyma-Taymyr ($F_{ST} = 0.098$) and Kolyma-Lena ($F_{ST} = 0.065$) show a separation that is about three to ten times stronger than between Taymyr-Lena ($F_{ST} = 0.010$) as indicated by the pairwise genetic difference analyses (Table 3S in the Electronic supplementary information). This is supported by the dbRDA result, which indicates that a significant proportion of the genetic

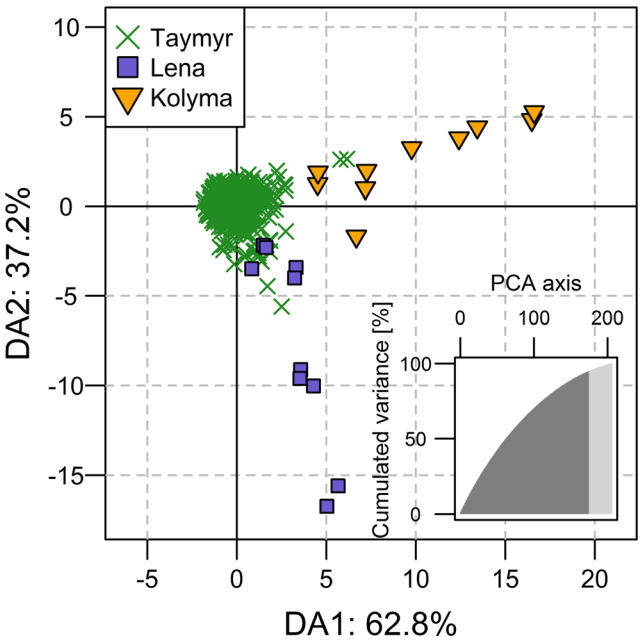

Fig. 2 Ordination based on allelic frequencies at eight microsatellite loci for individuals analysed from three regions. The inset gives the cumulative variance retained by the principal components analysis (PCA) axes for the discriminant analysis of principal components (DAPC)

variation (18.2%, $p < 0.001$) in the dataset is related to regional differences (Electronic supplementary information S4).

Differentiation among Taymyr Peninsula sub-populations

Pairwise genetic differences between Taymyr sub-populations show low mean values of $F_{ST} = 0.005$ (0.000–0.019; Table 3). Similar results are obtained for other differentiation measures (Nei's G_{ST}, Hedrik's G_{ST}, Jost's D; Electronic supplementary information S3). We find that 45 of the 55 (80%) pairwise F_{ST} comparisons have significant differences (above zero at $p < 0.05$).

Cluster methods failed to clearly differentiate individuals (Electronic supplementary information S5). However, some structure among the 743 individuals can be found, as dbRDA revealed that a significant proportion of the genetic variation (2.55%, $p < 0.001$) can be attributed to differentiation among sub-populations (Electronic supplementary information S4). In particular, the 'forest line' stands (TY04, CH17, TY09) are separated from the others and the explained variance is highest (0.11–0.54%), while it is low for 'single tree' stands and 'dense forest' stands (Fig. 3a).

The DAPC revealed the same pattern of more genetically distinct 'forest line' stands, as the individuals from these sites are placed on the edges of the ordination graph (Fig. 4). The group centroids of 'single tree' stands (TY02, CH12) and 'dense forest' stands (CH02, CH06) are mostly located around the origin of the plot of the first two DAPC axes, indicating that they lack strong separation from each other. In contrast, the group centres of two 'forest line' stands (TY09, CH17) have high axes values, indicating that each of them has a rather unique composition. Similar results are obtained when the structure is investigated on all three significant DAPC axes (not shown). To track either local recruitment or immigration of genotypes, we compared the position of group centres of recruits (individuals < 2 m tall) with the position of the respective mature trees (height > 2 m). In 'forest line' stands recruits are, in general, closer to the mature trees, while for 'dense forests' at the southernmost locations the strongest separation between these height classes is seen (Fig. 3b).

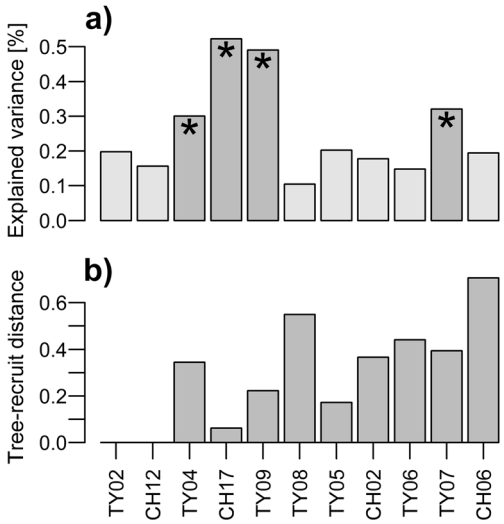

Fig. 3 **a** Percent variation explained of individual sub-populations compared with all other individuals in a distance-based redundancy analysis (dbRDA) and **b** inter-sub-population distance between group centres of trees (height > 2 m) and recruits (height < 2 m) for first and second discriminant functions (similar to axes in Fig. 4). Asterisks indicate significant values at $p < 0.001$. Includes both dbRDA and DAPC for all 743 individuals

Genetic diversity of sub-populations on the Taymyr peninsula

Out of the 743 tested individuals 75 are from 24 clonal groups of two to six individuals per group. These 24 clones are present only in 'single tree' stands, CH12 and TY02, which

Table 3 Pairwise Weir and Cockerham's F_{ST} values for sub-populations on the Taymyr Peninsula

	TY02	CH12	TY04	CH17	TY09	TY08	TY05	CH02	TY06	TY07	CH06
TY02											
CH12	0.005*										
TY04	0.001	0.007*									
CH17	0.007*	0.013*	0.012*								
TY09	0.005*	0.000	0.003*	0.012*							
TY08	−0.003	−0.002	0.003*	0.009*	0.003*						
TY05	0.004*	0.005*	0.002*	0.017*	0.003*	−0.002					
CH02	0.006*	−0.002	0.007*	0.013*	0.006*	−0.001	0.004*				
TY06	0.004*	0.002*	0.003*	0.018*	0.004*	0.003*	−0.001	0.002*			
TY07	0.001*	0.007*	0.002*	0.014*	0.004*	0.001	0.002*	0.006*	0.003*		
CH06	0.005*	0.011*	0.004*	0.019*	0.005*	0.003*	−0.004	0.003*	0.003*	0.003*	

*Significant values

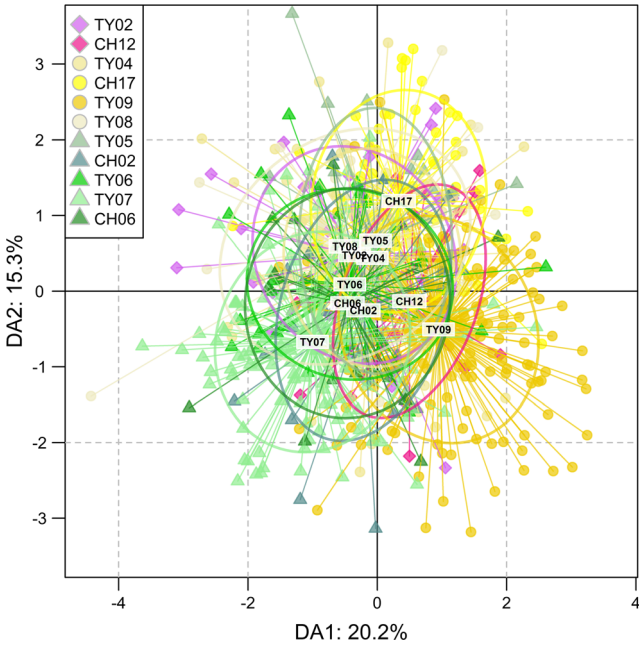

Fig. 4 Discriminant analysis of principle components (DAPC) on sub-population membership based on a principal components analysis (PCA) with Bruvo's genetic distances between pairwise individuals. The labels indicate the centres of individuals for a sub-population and the symbols refer to the stand type as in Fig. 1; first and second discriminant analysis (DA) functions together explain 35.5% of the variation

contain 11 (~ 39%) and 64 (~ 69%) clonal individuals, respectively. These samples were excluded from further analyses.

Sub-populations on the Taymyr Peninsula markedly differ in several allele-based diversity measures. We found 12–14 alleles for each of the 11 sub-populations, with lowest allele richness occurring in the 'forest line' stands TY04 and CH17 (Table 4). Each sub-population has 7–11 rare alleles and 1–4 private alleles, with low values mostly found in the 'forest line' stands. All sub-populations have a high expected heterozygosity (H_E = 0.826–0.893) and show signs of homozygote excess with varying deviation from the Hardy-Weinberg equilibrium of between 20 and 60% for all 8 loci. These differences from the Hardy-Weinberg equilibrium form a roughly unimodal shape with a maximum at 'forest line' stands.

Discussion

Larix population dynamics across the arctic treeline in Siberia

Strong genetic exchange across the treeline ecotone in Siberia

The sub-populations on the Taymyr Peninsula covering a 300-km-long transect show low genetic differentiation (F_{ST} = 0.005) in comparison with the observed interregional differences (Taymyr, Lena, Kolyma, F_{ST} = 0.042). However, the estimation of the F_{ST} values depends on the heterozygosity and the allele diversity and can thus lead to misinterpretations of small values (Jost 2008; Heller and Siegismund 2009; Meirmans and Hedrick 2011). We based our interpretation on the observed pairwise F_{ST} values, using dbRDA to test the sub-population differences, which explained only ~ 2.5% of the total variation. Our conclusion that the sub-populations have weak genetic differentiation is further supported from alternative measures (Hedrick's G'_{ST}, Jost'S D), which supply more robust estimates. We found similar patterns of pairwise genetic differences with values reaching ~ 6%. The F_{ST} values are low in comparison with other *Larix* studies that sampled a larger area (*L. gmelinii* = 0.017, *L. sibirica* = 0.071, *L. cajanderi* = 0.038; Oreshkova et al. 2013) and in

Table 4 Summary statistics for genetic diversity of the 11 sub-populations from the Taymyr Peninsula

Sub-population name	Stand type	Alleles	Rare alleles	Private alleles	H_E	HWE deviances (%)
TY02	'Single trees'	13.5 ± 4.3 ad	9.6 ± 2.4 bd	2.8 ± 1.6 bc	0.872 ± 0.061 d	30.75 ± 15.22 bc
CH12	'Single trees'	13.9 ± 5.0 b	10.2 ± 1.4 d	2.9 ± 1.0 bc	0.863 ± 0.094 a	37.50 ± 0.00 a
TY04	'Forest line'	13.1 ± 4.4 e	7.4 ± 2.3 a	2.7 ± 0.8 bc	0.877 ± 0.062 ef	33.25 ± 15.20 ab
CH17	'Forest line'	12.4 ± 3.9 c	7.4 ± 2.1 a	1.6 ± 1.2 de	0.826 ± 0.110 c	49.50 ± 14.96 d
TY09	'Forest line'	13.3 ± 4.2 ae	8.2 ± 2.7 ac	2.2 ± 1.3 bd	0.872 ± 0.081 d	50.13 ± 13.24 d
TY08*	'Forest line'	13.7 ± 5.1 f	8.6 ± 1.8 bc	1.0 ± 0.9 e	0.876 ± 0.070 f	44.88 ± 12.70 de
TY05*	'Dense forest'	13.6 ± 4.7 d	10.0 ± 1.9 d	1.7 ± 1.2 d	0.893 ± 0.058 g	38.75 ± 13.47 abe
CH02	'Dense forest'	13.3 ± 4.6 a	6.8 ± 1.7 a	3.8 ± 1.0 a	0.862 ± 0.106 a	37.63 ± 10.13 ab
TY06*	'Dense forest'	13.9 ± 4.5 b	10.9 ± 2.2 d	2.3 ± 1.1 b	0.893 ± 0.056 g	59.25 ± 13.49 f
TY07	'Dense forest'	13.3 ± 4.0 ae	8.6 ± 2.3 bc	3.1 ± 1.2 c	0.881 ± 0.060 be	39.63 ± 14.66 ae
CH06	'Dense forest'	13.4 ± 4.4 a	9.0 ± 1.7 bc	2.7 ± 1.3 bc	0.883 ± 0.069 b	26.50 ± 4.08 c

Number of rare alleles are alleles with < 1% overall frequency; HWE deviances are the percentage of counts significantly deviating from Hardy-Weinberg equilibrium under α = 0.05 for each locus individually. Letters a–g indicate sub-populations in which the mean values are not different at α = 0.01

comparison with those that sampled a smaller area (*L. gmelinii* = 0.117, *L. gmelinii* in Kamchatka = 0.039, *L. gmelinii* in Magadan = 0.026 with populations ~ 25 km apart from each other; Oreshkova et al. 2015). The weak genetic structure of sub-populations on the Taymyr Peninsula led also to the failure of statistical approaches to consistently cluster the data (similar to Nishimura and Setoguchi 2011). Additionally no regions could be identified with relatively more or less migration (cp. Petkova et al. 2016). Furthermore, we observed a higher allele richness and higher expected heterozygosity compared with results of the nuclear SSR marker study from Oreshkova et al. (2013) (*L. gmelinii*: $N_A = 8.1$, $H_E = 0.70$).

Our results indicate that the sub-populations of *Larix* are not well genetically differentiated and that high gene flow occurs between tree stands on the southern Taymyr Peninsula over at least c. 300 km. This contradicts previously published assumptions that *Larix* pollen (Niemeyer et al. 2015) and seeds (Kharuk et al. 2006), both representing vehicles for nuclear microsatellite genetic exchange, have poor dispersal abilities. High gene flow is common in wind-pollinated trees (Burczyk et al. 2004) such as *Quercus robur* L., where pollination can occur over distances of 80 km (Buschbom et al. 2011). It has also been demonstrated in a currently expanding population of *L. decidua* in the Alps, where no genetic erosion at the leading edge could be detected (Pluess 2011). Therefore, *Larix* might indeed be characterised by a high potential for long-distance dispersal, which enables rapidly moving species to minimise founder effects.

Slow movement of the leading edge

Despite the inferred strong genetic mixing, we found some significant but weak genetic structure among sub-populations on the Taymyr Peninsula. The tree stands at the 'forest line' differ in their allele patterns between each other and show the strongest deviance from all sub-populations (dbRDA). Furthermore, they slightly differ from southern 'dense forest' stands with respect to their low allelic richness (allele richness, private alleles, rare alleles). These sub-populations showed increased deviances from HWE, which might indicate inbreeding and is a suggestive of a currently slow moving front. This finding is in conjunction with the observed high recent recruitment at these locations (Wieczorek et al. 2017). The highest proportion of trees was recruited during the last decade at 'forest line' stands TY04 and TY09, and at CH17 a peak of recruits occurred ~ 25 years (approximately one generation) earlier. It is likely that only a few individuals colonised these sites causing a reduction of diversity compared with their source populations farther south. These stands have not yet been able to recover the state of a 'dense forest' at mutation-drift equilibrium because of restricted gene exchange over short distances as a result of a high proportion of local pollination and effective seed recruitment (Hartl and Clark 2007; Arenas et al. 2012). This scenario is corroborated by the observation of relatively few rare and private alleles at these 'forest line' stands and the strongest heterozygosity deficit, which is also considered a sign of local recruitment. In general, we find no evidence for any systematic influence on the inferences of the allele-based statistics introduced by the sampling scheme with unequal pairwise sample distances in each of the sub-populations. However, the potential error introduced by the different sample distances cannot be unequivocally ruled out of having any apparent influence in our analyses. In summary, we observe a treeline that is most probably expanding and showing typical signs of a slow moving front.

'Single tree' stands as relicts of a former forest extension

If, according to this 'slow moving front' scenario, 'single tree' stands growing within arctic tundra originated from far-distance dispersed seeds, they should show signs of founder effects, i.e. weak gene diversity and strong genetic differentiation. Our results do not support this scenario, but instead show signs that these stands are relicts from a former dense forest population. Evidence of high gene flow in the former forest and conservation of high genetic diversity in the small populations of the rapidly contracted forest area (Arenas et al. 2012) can be seen from the following deductions. If tree groups were locally reproducing, they would accumulate local differences (Hartl and Clark 2007), but we find them to have the weakest separation from all of the other sub-populations, in terms of low pairwise F_{ST}, not significantly explained variation in dbRDA, and their central position in the DAPC. Allele diversity was not lost despite the probable switch from sexual to asexual reproduction through clonal growth (ramets) in times of unfavourable growing conditions. This inference is corroborated by the observed high percentage of clonal individuals (this study) and the apparent lack of recruits at these sites (Wieczorek et al. 2017). Such an enduring life-history strategy makes it possible for this species to survive long periods under harsh environmental conditions and to alter their growth-form back to upright stems if conditions get better (personal observations; e.g. black spruce, Laberge et al. 2000; Wieczorek et al. 2017). This supports the hypothesis that larches may have survived in northerly areas during the last glacial maximum in suitable microsites (Polezhaeva et al. 2010). This is also a plausible explanation for the weak population structure observed over 12 degrees of latitude in *Picea obovata* along the Yenisei river (Chen et al. 2014). The alternative hypothesis, that individuals in 'single tree' stands were recruited during modern times by long-distance dispersal of seeds or pollen, seems rather unlikely because we would then expect to find a closer relationship with 'forest line' stands. Pluess (2011), for example, found

very weak genetic separation among *L. decidua* populations that had rapidly colonised a foreland following glacier retreat in the Alps. Local recruitment from seeds in 'single tree' stands at the Taymyr Peninsula is limited, as indicated by low seed quality and corroborated by a simulation study (Wieczorek et al. 2017). However, to clearly distinguish between seed- and pollen-mediated gene flow, further studies are necessary. In summary, the observed lack of recruitment and the presence of potentially very old, persistent individuals in the 'single tree' stands support the past connectivity of these stands, which are currently enduring sub-optimal conditions (cf. Louy et al. 2013). Hence, we infer that the 'single tree' stands are imprints of past treeline dynamics. A simulation study at this focal area might support this hypothesis. However, the low level of difference between sub-populations as probably caused by high gene flow either led to classical methods producing non-conclusive results or supporting only a few and/or weak separations (F_{ST}, PCA, dbRDA or DIYABC support only the simplest scenario). Additionally, the complex life-history of the studied larches made realistic and easy to implement forward simulations infeasible. To face these challenges, we are currently implementing genetics into an individually based larch vegetation simulation model (LAVESI, Kruse et al. 2016).

Probable past and present treeline dynamics

Pollen and macrofossil records suggest a very fast range expansion of *Larix* in Siberia during the late glacial and early Holocene (Kienel and Siegert 1999; MacDonald et al. 2000b; Andreev et al. 2002) facilitated by long-distance seed dispersal and/or nearby glacial refugia (Polezhaeva et al. 2010). The expansion maximum ~ 200 km north of the modern treeline was reached roughly 8000 years ago and a retreat began c. 3800 years BP (MacDonald et al. 2000b). As a consequence, a complex set of alleles formed in the distribution range of larches on the southern Taymyr Peninsula over these thousands of years.

Due to summer climate cooling since the mid-Holocene as a result of lower insolation, larch retreated rapidly southwards until c. 2000 years. ago in the vicinity of the 'single tree' stand CH12 and, since then, larch has been present but only at a very low population density (Klemm et al. 2016). Recent temperature increases in the Arctic mean that Holocene optimum temperatures will possibly be reached again during forthcoming decades (Arctic Climate Impact Assessment 2004; IPCC 2013). Forests on the Taymyr Peninsula are responding to this warming by infilling, especially at the 'forest line' (Wieczorek et al. 2017; Kharuk et al. 2006), but the rate of densification might lag by decades due to the long generation times of this species, a hypothesis corroborated by simulations (Kruse et al. 2016). The densification of forests will lead to a lower albedo, thereby further increasing local temperatures with the danger of a global feedback (Bonan 2008). This may be amplified by colonisation of tundra areas: however, our data suggest that this is a rather slow process and current forests are not immediately reacting in this way. Accordingly, vegetation-climate feedback will probably only be realised with a marked delay.

Conclusions

Our data on larch populations in north-central Siberia, generated by adapting a set of microsatellite markers originally designed for closely related species, reveal a high degree of connectivity among sub-populations of the treeline ecotone, spanning approximately 300 km in the lowlands of the southern Taymyr Peninsula. While the lack of differentiation among the populations suggests that dispersal is not limited within the area, our detailed analyses revealed that stands at the forest line are responding to current warming mainly by local infilling. They are characterised by a higher genetic distinctiveness and an increased degree of homozygosity. Equally, recruitment and reproduction of 'single tree' stands north of the current forest line is local. The krummholz larches growing in low numbers in the tundra show a (primarily) vegetative mode of reproduction and, thus, conserve their historical genetic composition. Due to the absence of recruitment from seeds, these sites are currently not participating in forest advancement into tundra areas. From these historical and contemporary population dynamics, we conclude that the latitudinal treeline is responding to recent climate warming with a distinct time-lag. Further studies, combining data from genetics, remote sensing, and simulations, and focusing on the forest line-tundra transgression zone should be conducted to verify our findings and reveal more detailed knowledge about probable future scenarios of treeline expansion.

Acknowledgements We thank our colleagues from joint Russian-German expeditions for support in the field. Special thanks to Alexey Kolmogorov and Bastian Niemeyer who helped collect samples during field work. We would like to thank Cathy Jenks for proofreading and improving the manuscript and two anonymous reviewers for their suggestions that helped to improve and clarify the manuscript.

Funding information The PhD positions of Stefan Kruse and Mareike Wieczorek were funded by the Initiating and Networking Fund of the Helmholtz Association and by the German Research Foundation, respectively, and Laura S. Epp acknowledges funding through the German Research Foundation grant EP98/2-1.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Data archiving statement Sampling locations, height classes and microsatellite genotype data are available at https://doi.pangaea.de/10.1594/PANGAEA.870947

R-scripts for the performed statistical analyses are stored online at https://doi.org/10.1594/PANGAEA.885765. These contain the preparation of the resampled datasets as well as the final data as R-objects; functions for reformatting the 'genind' objects for subsequent analyses in computer programs outside of the R-framework (CERVUS, STRUCTURE, ML-NULL, GENEPOP, MICRO-CHECKER, STRUCTURE, EEMS). The 'genind' objects are the formal class (S4) for individual genotypes in the R package 'adegenet'. In these objects, we included the population stratification—coding the region of sample origin, the sub-population name and the height class—for each individual sample.

References

Abaimov AP (2010) Geographical distribution and genetics of Siberian larch species. In: Osawa A, Zyryanova OA, Matsuura Y, Kajimoto T, Wein RW (eds) Permafrost Ecosystems—Siberian Larch Forests. Springer, Dordrecht, pp 41–58

Aizawa M, Yoshimaru H, Saito H et al (2009) Range-wide genetic structure in a north-east Asian spruce (*Picea jezoensis*) determined using nuclear microsatellite markers. J Biogeogr 36:996–1007

Andreev AA, Siegert C, Klimanov VA et al (2002) Late Pleistocene and Holocene vegetation and climate on the Taymyr Lowland, Northern Siberia. Quat Res 57:138–150

Arctic Climate Impact Assessment (2004) Impacts of a warming Arctic-Arctic climate impact assessment. Cambridge University Press, Cambridge

Arenas M, Ray N, Currat M, Excoffier L (2012) Consequences of range contractions and range shifts on molecular diversity. Mol Biol Evol 29(1):207–218. https://doi.org/10.1093/molbev/msr187

Ashley MV (2010) Plant parentage, pollination, and dispersal: how DNA microsatellites have altered the landscape. Crit Rev Plant Sci 29(3): 148–161. https://doi.org/10.1080/07352689.2010.481167

Bigelow NH, Brubaker LB, Edwards ME et al (2003) Climate change and Arctic ecosystems: 1. Vegetation changes north of 55° N between the last glacial maximum, mid-Holocene, and present. J Geophys Res 108:1–25

Bonan GB (2008) Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science 320(5882):1444–1449. https://doi.org/10.1126/science.1155121

Bruvo R, Michiels NK, D'Souza TG, Schulenburg H (2004) A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Mol Ecol 13(7):2101–2106. https://doi.org/10.1111/j.1365-294X.2004.02209.x

Burczyk J, DiFazio SP, Adams WT (2004) Gene flow in forest trees: how far do genes really travel? For Genet 11:179–192

Buschbom J, Yanbaev Y, Degen B (2011) Efficient long-distance gene flow into an isolated relict oak stand. J Hered 102:464–472

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol 24(3):621–631. https://doi.org/10.1093/molbev/msl191

Chen C, Liewlaksaneeyanawin C, Funda T et al (2009) Development and characterization of microsatellite loci in western larch (*Larix occidentalis* Nutt.) Mol Ecol Resour 9:843–845

Chen J, Tsuda Y, Stocks M, Källman T, Xu N, Kärkkäinen K, Huotari T, Semerikov VL, Vendramin GG, Lascoux M (2014) Clinal variation at phenology-related genes in spruce: parallel evolution in FTL2 and *Gigantea*? Genetics 197(3):1025–1038. https://doi.org/10.1534/genetics.114.163063

Chybicki IJ, Burczyk J (2008) Simultaneous estimation of null alleles and inbreeding coefficients. J Hered 100:106–113

Corlett RT, Westcott DA (2013) Will plant movements keep up with climate change? Trends Ecol Evol 28(8):482–488. https://doi.org/10.1016/j.tree.2013.04.003

Culley TM, Weller SG, Sakai AK, Putnam KA (2008) Characterization of microsatellite loci in the Hawaiian endemic shrub *Schiedea adamantis* (Caryophyllaceae) and amplification in related species and genera. Mol Ecol Resour 8(5):1081–1084. https://doi.org/10.1111/j.1755-0998.2008.02161.x

Davis MA, Douglas MR, Webb CT et al (2015) Nowhere to go but up: impacts of climate change on demographics of a short-range endemic (*Crotalus willardi obscurus*) in the sky-islands of southwestern North America. PLoS One 10:1–20

Harsch MA, Hulme PE, McGlone MS, Duncan RP (2009) Are treelines advancing? A global meta-analysis of treeline response to climate warming. Ecol Lett 12(10):1040–1049. https://doi.org/10.1111/j.1461-0248.2009.01355.x

Hartl DL, Clark AG (2007) Principles of population genetics. Sinauer Associates, Inc. Publishers, Sunderland

Hedrick PW (2005) A standardized genetic differentiation measure. Evol; Int J Organ Evol 59(8):1633–1638. https://doi.org/10.1111/j.0014-3820.2005.tb01814.x

Heller R, Siegismund HR (2009) Relationship between three measures of genetic differentiation GST, DEST and G'ST: how wrong have we been? Mol Ecol 18(10):2080–2083. https://doi.org/10.1111/j.1365-294X.2009.04185.x

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25(15):1965–1978. https://doi.org/10.1002/joc.1276

Holleley CE, Geerts PG (2009) Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. BioTechniques 46(7):511–517. https://doi.org/10.2144/000113156

Holtmeier F-K, Broll G (2005) Sensitivity and response of northern hemisphere altitudinal and polar treelines to environmental change at landscape and local scales. Glob Ecol Biogeogr 14(5):395–410. https://doi.org/10.1111/j.1466-822X.2005.00168.x

Holtmeier F-K, Broll G (2007) Treeline advance—driving processes and adverse factors. Landscape Online 1:1–33

IPCC (2013) Climate change 2013: The physical science basis. In: Stocker TF, Qin D, Plattner G-K et al (eds) Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK

Isoda K, Watanabe A (2006) Isolation and characterization of microsatellite loci from *Larix kaempferi*. Mol Ecol Notes 6(3):664–666. https://doi.org/10.1111/j.1471-8286.2006.01291.x

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart T, Ahmed I (2011) Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27(21):3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart T, Devillard S, Balloux F et al (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:1–15

Jost LOU (2008) G_{ST} and its relatives do not measure differentiation. Mol Ecol 17(18):4015–4026. https://doi.org/10.1111/j.1365-294X.2008.03887.x

Kalinowski ST, Taper ML (2006) Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. Conserv Genet 7(6): 991–995. https://doi.org/10.1007/s10592-006-9134-9

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. Peer J 2:e281. https://doi.org/10.7717/peerj.281

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6:1–10

Kharuk VI, Ranson KJ, Im ST, Naurzbaev MM (2006) Forest-tundra larch forests and climatic trends. Russ J Ecol 37(5):291–298. https://doi.org/10.1134/S1067413606050018

Kharuk VI, Ranson KJ, Im ST, Oskorbin PA, Dvinskaya ML, Ovchinnikov DV (2013) Tree-line structure and dynamics at the northern limit of the larch forest: Anabar plateau, Siberia, Russia. Arct Antarct Alp Res 45(4):526–537. https://doi.org/10.1657/1938-4246-45.4.526

Khasa DP, Newton CH, Rahman MH, Jaquish B, Dancik BP (2000) Isolation, characterization, and inheritance of microsatellite loci in alpine larch and western larch. Genome 43(3):439–448. https://doi.org/10.1139/g99-131

Kienel U, Siegert C (1999) Late quaternary palaeoenvironmental reconstructions from a permafrost sequence (North Siberian lowland, SE Taymyr Peninsula)—a multidisciplinary case study. Boreas 28(1): 181–193. https://doi.org/10.1111/j.1502-3885.1999.tb00213.x

Klemm J, Herzschuh U, Pestryakova LA (2016) Vegetation, climate and lake changes over the last 7000 years at the boreal treeline in north-central Siberia. Quat Sci Rev 147:422–434

Kruse S, Wieczorek M, Jeltsch F, Herzschuh U (2016) Treeline dynamics in Siberia under changing climates as inferred from an individual-based model for *Larix*. Ecol Model 338:101–121. https://doi.org/10.1016/j.ecolmodel.2016.08.003

Laberge M-J, Payette S, Bousquet J (2000) Life span and biomass allocation of stunted black spruce clones in the subarctic environment. J Ecol 88:584–593

Louy D, Habel JC, Abadjiev S, Schmitt T (2013) Genetic legacy from past panmixia: high genetic variability and low differentiation in disjunct populations of the eastern large heath butterfly. Biol J Linn Soc 110(2):281–290. https://doi.org/10.1111/bij.12144

MacDonald GM, Gervais BR, Snyder JA, Tarasov GA, Borisova OK (2000a) Radiocarbon dated *Pinus sylvestris* L. wood from beyond tree-line on the Kola Peninsula, Russia. The Holocene 10(1):143–147. https://doi.org/10.1191/095968300667807510

MacDonald GM, Velichko AA, Kremenetski CV, Borisova OK, Goleva AA, Andreev AA, Cwynar LC, Riding RT, Forman SL, Edwards TWD, Aravena R, Hammarlund D, Szeicz JM, Gattaulin VN (2000b) Holocene treeline history and climate change across northern Eurasia. Quat Res 53(03):302–311. https://doi.org/10.1006/qres.1999.2123

MacDonald G, Kremenetski KV, Beilman D (2008) Climate change and the northern Russian treeline zone. Philos Trans R Soc B: Biol Sci 363(1501):2283–2299. https://doi.org/10.1098/rstb.2007.2200

Marshall JC, Kingsbury BA, Minchella DJ (2009) Microsatellite variation, population structure, and bottlenecks in the threatened copperbelly water snake. Conserv Genet 10(2):465–476. https://doi.org/10.1007/s10592-008-9624-z

Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. Mol Ecol Resour 11(1):5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Micheneau C, Dauby G, Bourland N, Doucet J-L, Hardy OJ (2011) Development and characterization of microsatellite loci in *Pericopsis elata* (Fabaceae) using a cost-efficient approach. Am J Bot 98:e268–e270

Miller CR, Waits LP (2003) The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): implications for conservation. Proceedings of the National Academy of Sciences of the United States of America 100:4334–4339

Nei M (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America 70:3321–3323

Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. Ann Hum Genet 41(2):225–233. https://doi.org/10.1111/j.1469-1809.1977.tb01918.x

Niemeyer B, Herzschuh U, Pestryakova LA (2015) Vegetation and lake changes on the southern Taymyr Peninsula, northern Siberia, during the last 300 years inferred from pollen and *Pediastrum* green algae records. The Holocene 25:596–606

Nishimura M, Setoguchi H (2011) Homogeneous genetic structure and variation in tree architecture of *Larix kaempferi* along altitudinal gradients on Mt. Fuji. J Plant Res 124(2):253–263. https://doi.org/10.1007/s10265-010-0370-1

Oksanen J, Blanchet FG, Friendly M et al. (2016) Vegan: community ecology package

Oreshkova NV, Belokon MM, Jamiyansuren S (2013) Genetic diversity, population structure, and differentiation of Siberian larch, Gmelin larch, and Cajander larch on SSR-marker data. Russ J Genet 49(2): 178–186. https://doi.org/10.1134/S1022795412120095

Oreshkova NV, Sedel'nikova TS, Pimenov AV, Efremov SP (2014) Analysis of genetic structure and differentiation of the bog and dry land populations of *Pinus sibirica* du tour based on nuclear microsatellite loci. Russ J Genet 50(9):934–941. https://doi.org/10.1134/S1022795414090105

Oreshkova NV, Vetrova VP, Sinelnikova NV (2015) Genetic and phenotypic diversity of *Larix cajanderi* Mayr in the north of the Russian Far East. Contemp Probl Ecol 8(1):9–20. https://doi.org/10.1134/S1995425515010096

Park H, Kim Y, Kimball JS (2016) Widespread permafrost vulnerability and soil active layer increases over the high northern latitudes inferred from satellite remote sensing and process model assessments. Remote Sens Environ 175:349–358

Petkova D, Novembre J, Stephens M (2016) Visualizing spatial population structure with estimated effective migration surfaces. Nat Genet 48:94–100

Piotti A, Leonardi S, Piovani P, Scalfi M, Menozzi P (2009) Spruce colonization at treeline: where do those seeds come from? Heredity 103:136–145

Pluess AR (2011) Pursuing glacier retreat: genetic structure of a rapidly expanding *Larix decidua* population. Mol Ecol 20(3):473–485. https://doi.org/10.1111/j.1365-294X.2010.04972.x

Polezhaeva MA, Lascoux M, Semerikov VL (2010) Cytoplasmic DNA variation and biogeography of *Larix* Mill. in Northeast Asia. Mol Ecol 19:1239–1252

Polezhaeva MA, Semerikov VL, Pimenova EA (2013) Genetic diversity of larch at the north of Primorskii Krai and limits of *Larix olgensis* A. Henry distribution. Russ J Genet 49(5):497–502. https://doi.org/10.1134/S1022795413030149

R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna https://www.R-project.org/

Ray N, Excoffier L (2010) A first step towards inferring levels of long-distance dispersal during past expansions. Mol Ecol Resour 10(5): 902–914. https://doi.org/10.1111/j.1755-0998.2010.02881.x

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for windows and Linux. Mol Ecol Resour 8(1): 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments: a poor man's approach to genotyping for research and high-throughput diagnostics. Nat Biotechnol 18(2):233–234. https://doi.org/10.1038/72708

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett 9(5): 615–629. https://doi.org/10.1111/j.1461-0248.2006.00889.x

Semerikov VL, Lascoux M (2003) Nuclear and cytoplasmic variation within and between Eurasian *Larix* (Pinaceae). Am J Bot 90(8): 1113–1123. https://doi.org/10.3732/ajb.90.8.1113

Semerikov VL, Polezhaeva MA (2007) Mitochondrial DNA variation pattern in larches of eastern Siberia and the Far East. Russ J Genet 43(6):646–652. https://doi.org/10.1134/S1022795407060075

Semerikov VL, Semerikov LF, Lascoux M (1999) Intra- and interspecific allozyme variability in Eurasian *Larix* Mill. species. Heredity 82(2): 193–204. https://doi.org/10.1038/sj.hdy.6884710

Semerikov VL, Semerikova SA, Polezhaeva MA, Kosintsev PA, Lascoux M (2013) Southern montane populations did not contribute to the recolonization of West Siberian plain by Siberian larch (*Larix sibirica*): a range-wide analysis of cytoplasmic markers. Mol Ecol 22(19):4958–4971. https://doi.org/10.1111/mec.12433

Shaw PW, Pierce GJ, Boyle PR (1999) Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. Mol Ecol 8(3):407–417. https://doi.org/10.1046/j.1365-294X.1999.00588.x

Shinde D, Lai Y, Sun F, Arnheim N (2003) Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. Nucleic Acids Res 31(3): 974–980. https://doi.org/10.1093/nar/gkg178

Svendsen J (2004) Late quaternary ice sheet history of northern Eurasia. Quat Sci Rev 23(11-13):1229–1271. https://doi.org/10.1016/j.quascirev.2003.12.008

Tsuda Y, Chen J, Stocks M, Källman T, Sønstebø JH, Parducci L, Semerikov V, Sperisen C, Politov D, Ronkainen T, Väliranta M, Vendramin GG, Tollefsrud MM, Lascoux M (2016) The extent and meaning of hybridization and introgression between Siberian spruce (*Picea obovata*) and Norway spruce (*Picea abies*): cryptic refugia as stepping stones to the west? Mol Ecol 25(12):2773–2789. https://doi.org/10.1111/mec.13654

Tsuda Y, Semerikov V, Sebastiani F, Vendramin GG, Lascoux M (2017) Multispecies genetic structure and hybridization in the *Betula* genus across Eurasia. Mol Ecol 26(2):589–605. https://doi.org/10.1111/mec.13885

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(3):535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Wagner S, Gerber S, Petit RJ (2012) Two highly informative dinucleotide SSR multiplexes for the conifer *Larix decidua* (European larch). Mol Ecol Resour 12:717–725

Walker DA, Raynolds MK, Daniëls FJA, Einarsson E, Elvebakk A, Gould WA, Katenin AE, Kholod SS, Markon CJ, Melnikov ES, Moskalenko NG, Talbot SS, Yurtsev BA, The other members of the CAVM Team (2005) The circumpolar Arctic vegetation map. J Veg Sci 16(3):267–282. https://doi.org/10.1111/j.1654-1103.2005.tb02365.x

Wandeler P, Smith S, Morin PA, Pettifor RA, Funk SM (2003) Patterns of nuclear DNA degeneration over time—a case study in historic teeth samples. Mol Ecol 12(4):1087–1093. https://doi.org/10.1046/j.1365-294X.2003.01807.x

Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). Mol Ecol 7(11):1569–1573. https://doi.org/10.1046/j.1365-294x.1998.00477.x

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wessel P, Smith WHF (1996) A global, self-consistent, hierarchical, high-resolution shoreline database. J Geophys Res: Solid Earth 101(B4): 8741–8743. https://doi.org/10.1029/96JB00104

Wetterich S, Rudaya N, Tumskoy V, Andreev AA, Opel T, Schirrmeister L, Meyer H et al (2011) Last glacial maximum records in permafrost of the east Siberian Arctic. Quat Sci Rev 30(21-22):3139–3151. https://doi.org/10.1016/j.quascirev.2011.07.020

Wieczorek M, Kruse S, Epp LS, Kolmogorov A, Nikolaev AN, Heinrich I, Jeltsch F, Pestryakova LA, Zibulski R, Herzschuh U (2017) Dissimilar responses of larch stands in northern Siberia to increasing temperatures—a field and simulation based study. Ecology 98(9):2343–2355. https://doi.org/10.1002/ecy.1887

Zibulski R, Herzschuh U, Pestryakova LA (2016) Vegetation patterns along micro-relief and vegetation type transects in polygonal landscapes of the Siberian Arctic. J Veg Sci 27:377–386

Zyryanova OA, Abaimov AP, Daimaru H, Matsuura Y (2010) Floristic diversity and its geographical background in Central Siberia. In: Osawa A, Zyryanova OA, Matsuura Y, Kajimoto T, Wein RW (eds) Permafrost ecosystems—Siberian larch forests. Ecological studies. Springer Netherlands, Dordrecht, pp 17–39