

Logan SA, Chytry M, Wolff K. [Genetic diversity and demographic history of the Siberian lime \(*Tilia sibirica*\)](#). *Perspectives in Plant Ecology, Evolution and Systematics* 2018, 33, 9-17.

Copyright:

© 2018. This manuscript version is made available under the [CC-BY-NC-ND 4.0 license](#)

DOI link to article:

<https://doi.org/10.1016/j.ppees.2018.04.005>

Date deposited:

15/05/2018

Embargo release date:

05 May 2019



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](#)

Title: Genetic diversity and demographic history of the Siberian lime (*Tilia sibirica*).

Authors: Samuel A. Logan¹, Milan Chytrý² and Kirsten Wolff¹

Affiliation addresses:

¹ School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU.

² Department of Botany and Zoology, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic.

Abstract

Tilia sibirica Bayer (Siberian lime) is endemic to the low mountain systems north of the Altai Mountains of southern Siberia, approximately 1,000kms to the east of the natural range limit of its closest congeneric, the small-leaved lime (*Tilia cordata*).

Some consider the taxon to be a sub-species of *T. cordata*. This putative pre-Last Glacial Maximum (LGM) relict may have had a stronghold in a refugium of the Altai Mountains and survived various waves of fluctuating climatic changes that occurred in the region. With continued climatic changes expected, these hardy but isolated populations can be important sources for population expansion. To date, we do not know the genetic status or history of this forest tree.

This study uses standard population genetic and Approximate Bayesian Computation (ABC) analyses of microsatellite data to determine the genetic diversity and differentiation, clonal occurrence and date of divergence of the two lime taxa. The results show that *T. sibirica* and *T. cordata* are distinct biological units with significant genetic differentiation. The ABC analysis suggests a (Middle) Pleistocene divergence. We have revealed low within-population genetic diversity as well as high

levels of clonality in *T. sibirica*. The focus now should be on restoring and conserving these small and isolated relict populations.

Keywords: Approximate Bayesian Computation; Clonality; Genetic diversity; Microsatellites; Refugium; Siberia; *Tilia cordata*; Tree genetics

1. Introduction

Tilia sibirica Bayer (Siberian lime) is an endemic taxon of southern Siberia. The Flora of Siberia (Vlasova, 1996) considers it as a separate species but it is regarded by others as a subspecies, *T. cordata* subsp. *sibirica* (Fischer ex Bayer) Pigott, or variant, *T. cordata* var. *sibirica* (Fischer ex Bayer) Maximowicz (Pigott, 2012).

The two taxa do not occur in sympatry naturally and while their morphologies are broadly similar, there are subtle differences in their leaves, ovaries, and stigma (Pigott, 2012). Intense logging throughout the 20th century (Grubov, 1940; Reverdatto, 1925) disturbed the southern Siberian hemiboreal forests containing lime (Khlonov, 1965). Indeed, some small lime populations may have been lost due to human activity (Amelin and Blyakharchuk, 2016). Some tree species may be resilient to changes in their local landscape because they contain sufficient genetic diversity and appropriate mechanisms for successful propagation (Hamrick, 2004). Despite populations being relatively isolated and small, *T. sibirica* may still have the genetic diversity required for successful local restoration. However, as far as we are aware, no genetic analysis has yet been carried out on the Siberian lime, *T. sibirica*. Therefore, the question is, are these small, isolated populations losing potentially important genetic diversity and adaptive genetic variation?

Krylov (1891, 1902) and other Russian authors (e.g. Polozhii and Krapivkina 1985; Krapivkina 2009) consider *T. sibirica* as part of the Tertiary relic flora, based on phytogeographical considerations. This suggests that the ancestral lineage of the current lime taxon within the region may date back several millions of years. *Tilia* records are frequently mentioned in Siberian palynological studies from the Tertiary but also in various periods of the Pleistocene (Bolikhovskaya and Shunkov, 2014; Khlonov, 1965). The genus is also well represented in the Siberian pollen record in

the period from the last interglacial (125 ka BP) to before the Last Glacial Maximum (LGM) which ended ~18–19 ka BP (Bolikhovskaya and Shunkov, 2014; Markova et al., 2009; Tarasov et al., 2005), although some of these records have been suspected to originate from redeposited Tertiary sediments (Amelin and Blyakharchuk, 2016). While much of the *Tilia* pollen is found between 60° and 65°N, a single *Tilia* pollen grain has been found as far north as the Laptev Sea (74°N), in deposits dated 7.2 ka BP (Naidina and Bauch, 2001), which corresponds to the climatic optimum of the Holocene. While long distance dispersal cannot be ruled out, *Tilia* pollen is poorly dispersed by wind, because it is usually pollinated by insects (Pigott, 2012), and these fossil finds consequently indicate that members of the genus may have once occurred further north than where they are found today.

With the expectation of a continued changing climate and further range shifts (Walther et al., 2002), it is important to understand the dynamics of temperate forest trees across their range and the genetic status of populations therein. To infer the genetic diversity, differentiation and clonality of *T. sibirica*, *Tilia*-specific microsatellite markers (Phuekvilai and Wolff, 2013) were used. These markers revealed high genetic diversity and population structure in *T. cordata* and *T. platyphyllos* in the UK (Logan et al., 2015) and Central Europe (Phuekvilai, 2014) and were able to identify clonal occurrence and genetic relatedness of European *Tilia* (Logan et al., submitted) .

We tested four hypotheses: [1] *T. cordata* and *T. sibirica* are two distinct taxa showing (significant) genetic differentiation from each other; [2] the divergence of the two taxa from a common ancestor predates the LGM; [3] given its limited and fragmented geographic range, *T. sibirica* populations show lower genetic diversity

than *T. cordata* populations; and [4] asexual reproduction is more common in *T. sibirica* than in *T. cordata*. Answering these fundamental questions will greatly advance our understanding of a potential Siberian relict and could indicate sites in need of conservation.

2. Material and Methods

2.1. Study sites and sample collection

We sampled six *Tilia sibirica* populations for genetic analyses from forests near the village of Kuzedeevo, Kemerovo Region (southern Siberia, Russia), which has the largest concentration of *T. sibirica* globally, while other populations are small, occurring in areas generally not exceeding 12ha (Amelin and Blyakharchuk, 2016; Khlonov, 1965). For comparison, we sampled *Tilia cordata* including two populations from the Vagay area in the Tyumen Region (West Siberian Plain, Russia), three populations from Białowieża, Poland and one from Stams, Austria (Table 1, Fig. 1). We sampled one leaf per tree from trees that were at least 10m apart. Details of related sampling for an ecological study at the Kuzedeevo site are described in Novák et al. (2014).

2.2. DNA extraction and amplification

Genomic DNA was extracted from leaf tissue using the CTAB (Doyle, 1992) method with slight modifications. A multiplex Polymerase Chain Reaction (PCR) procedure was carried out to amplify twelve microsatellite regions (Phuekvilai and Wolff, 2013). PCR conditions and parameters were as described in Phuekvilai and Wolff (2013). Microsatellites or Simple Sequence Repeats (SSRs) were genotyped using an ABI 3130XL Genetic Analyser (Applied Biosystems), and scored using Genemapper

(Applied Biosystems). Microsatellite fragments were binned manually and checked for inconsistencies.

2.3. Data analyses

Samples were checked for deviations from Hardy-Weinberg Equilibrium (HWE) using GENEPOP on the web v4.2 (Raymond and Rousset, 1995; Rousset, 2008) and tested for scoring errors, stuttering and large allele dropout in MICRO-CHECKER (Van Oosterhout et al., 2004). Null allele frequencies were estimated per locus using FreeNA (Chapuis and Estoup, 2007).

The presence of clones, *i.e.* individuals with identical multi-locus genotypes (MLG) within *T. sibirica* and *T. cordata* was identified in GenClone v2.0 (Arnaud-Haond and Belkhir, 2007), where a resampling procedure (1000 permutations in this study), was implemented. Individuals with missing data were removed prior to clonal analysis. Genotypic richness $R = (G-1/N-1)$, where G is the number of genotypes and N is the number of individuals, was estimated for all sampled *T. sibirica* and *T. cordata* populations. R will always show '0' when stands consist of a single clone and '1' when all sampled individuals are separate genets (Dorken and Eckert, 2001). To further describe clonal heterogeneity, an adapted estimate of the Simpson's complement index (*i.e.* D^*), independent of sample size (Pielou, 1969), was estimated in GenClone. The Simpson index (D) represents the probability that two randomly sampled plant individuals will belong to the same species (Simpson, 1949) and is widely used in ecology. The Simpson's complement index ($1-D$) of diversity is commonly reported in clonal studies as D^* (Arnaud-Haond et al., 2007) and ranges from '0' when all individuals within a population are clonal to '1' when all individuals are unique.

Additionally, the statistics P_{gen} , the probability of samples having the same MLG by chance following the methods of Parks and Werth (1993) and P_{sex} , the probability that a repeated MLG originates from sexual reproduction at the first re-encounter, were estimated. P_{gen} assumes populations are in Hardy-Weinberg Equilibrium (HWE). An adjusted measure $P_{gen}(f)$, which takes into account HW departure can also be estimated, providing a more conservative estimate of P_{sex} (Arnaud-Haond et al., 2007). Following clonal analyses, multiple copies of genotypes were removed during further analyses meaning only single genotypes were used.

Expected heterozygosity (H_E), total number of alleles (A), number of private alleles (A_p) and average number of alleles (N_A) were calculated in GenAEx 6.5 (Peakall and Smouse, 2012). Allelic richness (A_R), based on a minimum of six diploid individuals, was calculated in FSTAT v2.9.3.2 (Goudet, 1995, 2001) for all populations.

Genetic differentiation between the two taxa was estimated using a number of different diversity measures *i.e.* F_{ST} (Weir and Cockerham, 1984), D_{est} (Jost, 2008), $G_{ST_{est}}$ (Nei and Chesser, 1983), and $G'_{ST_{est}}$ (Hedrick, 2005), in Genepop on the web v4.2 and SMOGD v1.2.5 (Crawford, 2010), respectively. D_{est} is an estimate of actual differentiation, $G_{ST_{est}}$ is a nearly unbiased estimate of relative differentiation, and $G'_{ST_{est}}$ is an estimate of standardised genetic differentiation. These were calculated using 1000 bootstrap replicates which generated 95% confidence intervals.

Population differentiation (pairwise F_{ST} and their significance) was calculated in FSTAT v2.9.3.2. Allele frequencies adjusted for null alleles were also used to calculate expected heterozygosity and F_{ST} (without significance level – both sets of values are reported). To determine the distribution of genetic variation – among species, among populations within species, and within populations – we used an

analysis of molecular variance (AMOVA) implemented in Arlequin v3.5.1.3 (Excoffier and Lischer, 2010).

To test whether *T. cordata* and *T. sibirica* individuals form distinct genetic groups, a Principal Coordinates Analysis (PCoA) based on individual pairwise genetic distance, was performed in GenAlEx. In addition, the microsatellite data were pooled with no *a priori* information regarding population structure and a Bayesian genetic clustering analysis performed in the program STRUCTURE v2.3.4 (Falush et al., 2003; Pritchard et al., 2000). STRUCTURE assigns individuals to a predefined number of clusters (K), based on allele frequencies at each locus. In this case we set K to range from 1 to 7 (*T. sibirica*, *T. cordata* from Siberia, Poland and Austria plus three). STRUCTURE parameters were kept at the default settings, with a burn-in of 10^4 and MCMC iterations of 10^5 . Each run was replicated ten times. Model selection relied on the Evanno ΔK statistic (Evanno et al., 2005) estimated in STRUCTURE HARVESTER (Earl and vonHoldt, 2012). Assignment probabilities for the optimum K were averaged across runs using CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007). To visualise the data we used the program DISTRUCT v1.1 (Rosenberg, 2004).

2.4. Historic model

An Approximate Bayesian Computation (ABC) approach was implemented in DIYABC v2.1.0 (Cornuet et al., 2014; Cornuet et al., 2008) to infer the demographic history of *Tilia* in Siberia. Effective population size (N_e) and the time of divergence (td) were estimated for *T. cordata* and *T. sibirica*. The program uses a suite of summary statistics to test genetic data based on a set of user-defined historical parameters.

We tested a simple split model (SSM, Fig. 2), by setting four historical parameters (Table 2). The effective population sizes ($N1$ and $N2$) for each taxon ranged from 10 to 20,000. One condition was set regarding N_e of *Tilia sibirica* ($N1$) and *T. cordata* ($N2$), that is, both were smaller than or equal to the ancestral population size (N_a). Time of divergence (td), was set as minimum 10 generations to maximum 10,000 generations. *Tilia* can first flower when 12 – 40 years old and trees can live for >450 years (Pigott, 2012). With this in mind we used a conservative generation time of 100 years. Assuming a 100 year generation time, td translates to 1 thousand years (ka) – 1 million years (ma) before present (BP). A Generalised Stepwise Mutation (GSM) model was assumed and a mean microsatellite mutation rate of 10^{-5} to 10^{-3} was drawn from a uniform distribution. The maximum value for the mean coefficient (P) was changed from the default settings to 9.99E-01 as was the individual locus coefficient. These settings were changed because on earlier analyses, the curve produced from the DIYABC output never reached its peak (data not shown). All other default settings remained in place.

An option within the DIYABC program (pre-evaluate scenario prior combinations) allows scenarios and priors to be initially tested. A Principal Component Analysis (PCA) is performed on 100,000 simulated data sets and a method that ranks each statistic of the observed data set against that of the simulated data. If the simulated data are significantly different from the observed then that set of summary statistics may not be best suited in describing the priors for the scenario (Estoup et al., 2015). To determine this, we set the program to run analyses using all available summary statistics to determine an 'optimum set', on the grounds that the statistics that showed highly significant differences between the simulated datasets and the observed were to be omitted.

In total, 16 summary statistics were recorded to build the SSM reference table. One-sample statistics included mean number of alleles (A), genic diversity (H - Nei, 1987), allele size variance across loci (V), and mean Garza-Williamson's M (MGW). Two-sample statistics included the mean number of alleles ($A2P$), mean genic diversity ($H2P$), allele size variance ($V2P$), F_{ST} (Weir and Cockerham, 1984), Classification index (LIK), shared allele distances (DAS) and $(\delta\mu)^2$ distance between two samples (Goldstein et al., 1995). A total of 1×10^6 datasets were simulated. The option 'model check' was used to assess the goodness-of-fit. The analysis is similar to the 'pre-evaluate option', in that it simulates data sets and compares those to the observed data. However, on this occasion the posterior distributions of parameters are used rather than the prior distributions. As before, a PCA (this time performed on 1×10^6 data sets) and a ranked approach are used. The goodness-of-fit PCA graph presented visualises 10,000 (1%) simulated data sets.

3. Results

In total, 146 *Tilia* individuals from Siberia were genetically analysed (113 *T. sibirica* and 33 *T. cordata*). Additionally 42 *T. cordata* individuals from Poland and 22 from Austria were analysed and these data (following the removal of repeated MLGs) were used to compare summary statistics with the Siberian *Tilia*. MICRO-CHECKER v2.23 revealed no evidence of scoring error due to stuttering and no large allele dropout. There was some evidence of homozygote excess suggesting possible deviation from Hardy-Weinberg Equilibrium (HWE) but these were observed at different loci and were not consistent across populations.

3.1. Clonality in *T. sibirica*

From the 113 *T. sibirica* individuals sampled, 49 were identified as clones (43.4%). In total, there were 75 genotypes observed within the taxon, eleven of which (14.7%) were clonal. Clone size ranged from 2 to 11 trees sampled per population (location) with no identical genotypes shared across locations. Genotypic richness (R) for all sampled *T. sibirica* individuals was 0.661 (mean over all populations was 0.601). Simpson's D^* for all individuals was 0.978 (mean 0.819). These values were generally lower than for all sampled *T. cordata* individuals (Table 3). The probability of identical genotypes arising by chance (P_{gen}) and the probability that those repeated genotypes (at the first re-encounter) were the result of sexual reproduction (P_{sex}) was calculated for the entire *T. sibirica* and *T. cordata* datasets. As some loci deviated from HWE, we used the $P_{gen}(f)$ and $P_{sex}(f)$ options in the program. Both sets of values were <0.01 for all MLGs (Appendix A), suggesting that repeated genotypes were clones and not due to chance.

3.2. Genetic diversity and differentiation

Two groups were identified following the PCoA analysis based on pairwise genetic distance (Fig. 3). The individuals with orange symbols represent *T. sibirica* while the individuals with green symbols unite the *T. cordata* populations from the Vagay (West Siberian Plain), Białowieża (Eastern-Central Europe), and Stams (Western-Central Europe). Two groups ($K=2$) were also observed following STRUCTURE analysis (Fig. 4) and using Evanno's ΔK method (Appendix B1). Based on the molecular markers used in this study, there is a clear genetic divide of the two taxa which correlates with their respective geographic divide, *i.e.* *Tilia cordata* occurring in Europe and West Siberian Plain and *T. sibirica* further east in southern Siberia (see

Fig. 1). We also applied a STRUCTURE analysis using only the *T. sibirica* data set (parameters as above with K ranging from 1 – 9). Evanno's ΔK method suggested that sub-structure was evident ($K = 2$). However, the two genetic groups were mixed suggesting possible ancestral isolation and then a subsequent admixture event (see Appendices B2 & B3).

The microsatellite markers revealed the level of polymorphism within the *Tilia* samples. It is clear that *T. sibirica* has lower genetic diversity than *T. cordata* from the three different regions, based on the markers used in this study, as revealed by expected heterozygosity (H_E) and adjusted expected heterozygosity (H_{E_NULL}). Values ranged from 0.274 to 0.379 in *T. sibirica* and were significantly lower (Mann-Whitney U test, $W = 30$, $P = 0.004$) than in *T. cordata*, ranging from 0.523 to 0.609. Values were also significantly different (Mann-Whitney U test, $W = 30$, $P = 0.004$) when adjusted for null alleles (Table 4). Estimates of null allele frequencies for each locus within each population are provided in Appendix C. The average number of alleles (N_A) in *T. cordata* ranged from 4.92 to 5.83, being significantly higher (Mann-Whitney U test, $W = 30$, $P = 0.008$) than in *T. sibirica*, which ranged from 2.00 to 3.08. Allelic richness (A_R) in *T. cordata* ranged from 3.78 to 3.98 and was significantly higher (Mann-Whitney U test, $W = 30$, $P = 0.004$) than in *T. sibirica* where A_R was not higher than 2.38. The Fixation index (F_{IS}) ranged from -0.194 to 0.448 in *T. sibirica* and were significantly different from zero ($P < 0.05$) in two populations. F_{IS} values within *T. cordata* ranged from -0.047 to 0.158 (Table 4).

In total, 136 alleles were identified between the two taxa. The total number of alleles per locus ranged from 1 (*Tc8*) to 29 (*Tc963*). There were more private alleles found in *T. cordata* (81) than in *T. sibirica* (14). Genetic differentiation

between the two taxa varied depending on which measure was applied (Table 5). Locus *Tc920* showed the highest differentiation across all measures and may be useful as a species identifier. In *T. sibirica* allele size at this locus was 223bp and 230bp with a rare allele of 218bp, while allele size ranged from 221 to 242bp in *T. cordata*. Significant population pairwise F_{ST} – between the populations – was calculated in GENEPOP (Table 6) and ranged from 0.002 (populations RV25 and RV20) to 0.303 (populations RV25 and RK29). These values changed only slightly following adjustment for null alleles (0.005 and 0.286, Appendix D). There was high genetic differentiation between the two taxa revealed by an AMOVA (15.86% of the total variation, $P < 0.001$). Among populations within taxa variation was 6.83% ($P < 0.001$), while the remaining variation (77.31%, $P < 0.001$) was found within populations (Table 7).

3.3. *Tilia sibirica* and *T. cordata* Simple Split Model (SSM)

Following the ‘pre-evaluate scenario and priors’ option with all available summary statistics, the PCA graph – which visualises the pattern of variation within the data – clearly showed that the observed data set (yellow dot) was placed within the 1,000 (1%) simulated data sets (green dots, Appendix E). This suggests that these statistics were suited for the priors and fits the model. All statistics were close enough to the observed data (Appendix F). to accept the parameters and scenario priors (Estoup et al., 2015). The SSM (Fig.2) revealed a lower effective population size for *T. sibirica* ($N1$) than for *T. cordata* ($N2$). Modes (and 95% CI) values (Appendix G) were estimated to be 4760 (1,770 – 10,900) and 13,700 (6,480 – 18,000), respectively. The time of divergence (td), was estimated at 4,470 (1,540 – 9,010) generations ago. Considering a generation time of 100 years (Pigott, 2012),

this infers a split occurred around 447 (154 – 901) ka BP. Model checking assessed the goodness-of-fit and produced a PCA graph showing a large cluster of simulated data from the prior and a smaller cluster of data from the posterior predictive distribution with the observed data set placed within both (Fig. 5). Posterior checking using the ranked approach (Appendix H) suggests that the observed data fits the scenario well (Estoup et al., 2015).

4. Discussion

Genetic analyses of *Tilia sibirica* and *T. cordata* suggest that the two taxa are distinct biological units. This is apparent from the PCoA and STRUCTURE analyses (Figs. 3 and 4) and from strong genetic differentiation (Table 6). In some populations, this genetic differentiation was comparable to pairwise F_{ST} values between *T. cordata* and *T. platyphyllos* (Logan et al., 2015). The estimated divergence time obtained from the DIYABC analysis suggests that the two taxa diverged from a common ancestor before the LGM. Given the level of genetic difference between the two taxa and the ecological and geographical isolation of *T. sibirica* in southern Siberia, a (sub)species level split approximately 1540 - 9010 generations ago, *i.e.* in the Middle Pleistocene, sometime between 154 and 901 ka BP (based on a generation time of 100 years), seems plausible.

4.1. Low genetic diversity and structure in *Tilia sibirica*

Not surprisingly, *Tilia cordata* showed a higher degree of genetic diversity than *T. sibirica* (Table 4). This may be due to *T. cordata* having a larger population size across a broader geographic range and having generally less isolated populations than *T. sibirica*. Our results showed that *T. cordata* from Russia, Poland, and Austria

has slightly higher genetic diversity than UK populations (Logan et al., 2015). *Tilia cordata* is a diploid species (Logan et al., 2015; Pigott, 2012), and we can confirm that *T. sibirica* is also diploid because all individuals analysed showed a maximum of two alleles per locus.

Stochastic genetic drift can lower genetic diversity of rare species and small populations (Ellstrand and Elam, 1993) and this could be a reason for the low levels of expected heterozygosity seen here in *T. sibirica*. An alternative (or complementary) reason may be that *T. sibirica* is not an obligatory outcrossing species. However, only two of the five populations showed F_{IS} values significantly higher than zero (Table 4) and there is no evidence of selfing in other *Tilia* species. The set of microsatellite markers used in this study were designed from *T. platyphyllos* individuals (Phuekvilai and Wolff, 2013), so we cannot completely rule out ascertainment bias (Ellegren et al., 1995) as a reason for the observed low diversity in *T. sibirica*. However, we compared allele size range and the signature of the repeat motifs at two highly polymorphic loci in both *T. sibirica* and *T. cordata* to ensure that the microsatellite actually exists in *T. sibirica*. Both allele size range and motif signatures were very similar (data not shown). We are therefore confident that the lower genetic diversity we observed in *T. sibirica* is not solely attributed to ascertainment bias but is a result of isolation and fragmentation.

Despite no obvious sub-structuring in *T. sibirica* following the model selection from STRUCTURE HARVESTER (Fig. 4 and Appendix B1), when analysed using only the *T. sibirica* dataset, Evanno's ΔK method suggested that optimum $K = 2$ (Appendix B2). While this implies some level of genetic sub-structuring within *T. sibirica*, it appears that the two genetic groups are generally mixed. Janes et al. (2017) suggest that caution should be taken when considering optimum $K = 2$. In this instance, there

is no biological meaning in the apparent sub-structure of $K = 2$ within *T. sibirica*. When considering the STRUCTURE output for K (data not shown), rather than five genetic groups are suggested and indeed there is a small peak at $K = 5$ (Appendix B2). Once again the genetic groups appear mixed across all populations (Appendix B4). As part of any future conservation or restoration effort, genetic diversity of *T. sibirica* must be taken into consideration. Knowing those populations that are genetically differentiated and those that are genetically similar is a basis for conservation projects (Pautasso, 2009). Further analyses could be carried out using more populations from across the entire range of the Siberian larch. This may aid in any future management of genetic stock regarding re-stocking less diverse areas. Despite having relatively low within-population genetic diversity the taxon has clearly survived various waves of climatic fluctuations and while this may be partly due to its ability to propagate clonally, there is certainly still scope to initiate a restoration project using these potentially adaptive, native trees by replanting into areas where they are currently found and into those areas where they once were prior to logging.

4.2. Pleistocene split between *T. cordata* and *T. sibirica*

Assuming a generation time of 100 years and considering the summary statistics, the parameters set and the microsatellite markers used, we can infer a Middle Pleistocene split between *T. cordata* and *T. sibirica* around 4470 generations ago (approximately 447 ka BP). Considering some of the earliest evidence of putative *Tilia* species have been dated from the Tertiary (Pigott, 2012; Wolfe and Wehr, 1987) and a recently approximated age of the Tilioideae was inferred to be 17 million years (Richardson et al., 2015), the estimated divergence date from this study may

seem relatively recent. However, DIYABC analyses provided credible intervals – the highest value being 901 ka BP, corresponding to the Middle Pleistocene (Appendix G). A coalescent based study using nuclear loci from two closely related North American tree species, *Populus balsamifera* and *P. trichocarpa*, inferred a divergence date of ~75 ka BP (Levsen et al., 2012) and while this is much more recent than the Early Pleistocene split of *Quercus* spp. (Bagnoli et al., 2015), it does suggest that climatic fluctuations of the Middle to Late Pleistocene had a strong effect on tree species causing lineage divergence. Regular expansion and retraction of *Tilia* populations in the Altai region (Bolikhovskaya and Shunkov, 2014) and climatic fluctuations (Groisman et al., 2013), may have strongly influenced the genetic structure of the two taxa leading to genetic isolation and difference which may have subsequently led to a Middle/Late Pleistocene split. A recent study based on the analysis of the glacial geomorphology in the Altai region suggests that only the higher altitudes in the central part of this mountain system were glaciated in the LGM, while low-altitude areas in the north with current occurrences of *Tilia sibirica* were ice-free (Blomdin et al., 2016). A palaeovegetation model suggested that non-glaciated lower mountainous landscapes of the north-eastern Altai were suitable habitats for open woodland vegetation even during the LGM (Hais et al., 2015). Although the most probable vegetation of this area in the LGM was open larch and sub-alpine coniferous woodland, persistence of small populations of clonally growing *Tilia* at topographically sheltered sites with suitable mesoclimate cannot be ruled out even for the LGM period.

We have revealed considerable clonal reproduction (42.5%) within *T. sibirica* (Table 3). Clonal growth in *T. sibirica* was also observed in the field by Novák et al. (2014). While two of the five *T. sibirica* populations from this present study showed

negative F_{IS} values, these were not significantly different from zero (Table 4). The level of clonality observed here in the *T. sibirica* was greater than that observed in *cordata* from Russia, Poland and Austria (Table 3). This may be the result of difficult generative regeneration due to competition of *Tilia* seedlings from tall forbs at the southern Siberian site. As this site receives high precipitation and has fertile soils, tall forbs spread quickly in all open areas including canopy gaps and form a dense herb layer, in which it is difficult for tree seedlings to survive (Novák et al., 2014). Clonality in *T. sibirica* is also greater than the high level of clonal reproduction (25% of the samples analysed) noted in UK *T. cordata* and *T. platyphyllos* (Logan et al., 2015). A more comprehensive study of clonal reproduction in *Tilia* populations across their entire European range is required to better understand its role in the genus.

4.3. *Tilia* relicts from Siberian refugia?

Phuekvilai (2014) suggested a possible European colonization route for *Tilia cordata* from a more easterly refugium. Putative refugia from the Caucasus, Caspian Sea region (Hewitt, 1999) and the central Russian Plains (Markova et al., 2009; Svenning et al., 2008) have been documented. Possible cryptic refugia at northern latitudes (northern refugia hypothesis) have been reviewed and tested (Bhagwat and Willis, 2008; Normand et al., 2011; Stewart and Lister, 2001; Svenning et al., 2008; Väiliranta et al., 2011; Willis et al., 2000; Willis and Van Andel, 2004; Willis and Whittaker, 2000). This led to many authors suggesting that the previous paradigm of largely treeless full glacial Central Europe should be revised, although Tzedakis et al. (2013) could not find support for the existence of refugia of temperate trees.

The SSM divergence time of 447 ka BP (Appendix G) suggests *Tilia* survived in a LGM refugium at northern latitudes. There is some evidence suggesting *T. cordata* surviving at higher latitudes during the maximum extent of ice sheets. Svenning et al. (2008) modelled the LGM distribution of several boreal and nemoral tree species and concluded that while nemoral species were largely confined to traditional southern refugia, *T. cordata* and a few others may have found suitable conditions further north. Bhagwat and Willis (2008) suggested that species with a current northern range of more than 60°N may have been able to survive at northern latitudes during the LGM. *Tilia cordata* has a northern geographical range of 65°N in Norway and Finland and 63°N in Russia. The species can survive undamaged at temperatures as low as -48°C and can clonally reproduce through vegetative spread (Pigott, 2012). Our study has revealed that *T. sibirica* may also invest more resources on reproducing asexually in less optimum conditions. Bhagwat and Willis (2008) suggest that having the ability to vegetatively propagate in extreme environments can prolong the existence of woody species at northern locations. Under suboptimum conditions and in the absence of sexual reproduction, it is likely that *T. cordata* and *T. sibirica* persisted in northern refugia during the glacial periods and clonal propagation played an important role in their survival, and still does in *T. sibirica* growing in a continental area with long and cold winters.

4.4 Implications for conservation

Protecting habitats and associated species is an essential part of wildlife conservation and today we have many ways to identify threatened taxa and landscapes. Identifying species with low genetic diversity and genetically isolated populations is an important part of that conservation and subsequent restoration

process. This can lead to species' being awarded long-term protection. However, protecting individual species may not necessarily lead to protecting habitat (Bragina et al., 2015). The Altai Mountains and many other forested areas within Siberia have had a long history of fragmentation from both legal and illegal logging despite having protected status (Shchur et al., 2017) although this 'fragmentation' rarely leads to the total clearance of Siberian forest unlike the outcome of many ancient European forests. That said, Amelin and Blyakharchuk (2016) point out that some of the 'naturally fragmented' populations of *T. sibirica* have in the past been lost due to local human activity. The concern here is that the current lime populations, which may have existed in isolation for centuries if not millennia, may still contain adaptive genetic potential and so any further loss will be hugely detrimental.

While it would be advantageous to have sampled and analysed more populations from across the entire range of *T. sibirica* to get a global estimate of genetic diversity within the taxon, we have revealed three populations from the six analysed that could be used as potential stock for assisted restoration and conservation. Based on our results, populations RK12, RK28 and RK29 consist of fewer clones and are genetically similar to the other three populations (RK21, RK22, and RK38 which are ~70% clonal). Individual trees from the RK12, RK28 and RK29 sites could be used to restock the other sites. Furthermore, RK12 has slightly higher genetic diversity and so could be used selectively to restore populations in forests near the Kuzedeevo village. We acknowledge that further genetic analyses from a wider cohort of individuals should be carried out to fully understand their conservation requirements. That said, we believe this study, the first to genetically analyse *T. sibirica*, will encourage just that. In the current geographic range of *T. sibirica*, wet climate and rich soils support rapid

development of tall forbs in forest canopy openings, which disadvantages recruitment of tree seedlings but benefits vegetatively regenerating woody plants (Lashchinskiy, 2009; Novák et al., 2014). We recommend that an *ex situ* approach be taken where *T. sibirica* trees are grown elsewhere then replanted back into those areas where they currently exist and possibly even extended to areas where they once grew prior to being lost to localised logging. Furthermore, providing some protected status may also benefit these small isolated populations. Currently, a part of the forest near Kuzedeevo is protected as a State Nature Monument 'Lipovy Ostrov' (Lime Island, Amelin and Blyakharchuk, 2016). We encourage additional 'lime islands' to be established as State Nature Monuments in an attempt to ensure the longevity of this ancient Siberian relict across its natural range.

5. Conclusion

This is the first genetic study carried out on the Siberian lime (*Tilia sibirica*), which is endemic to a small number of sites in southern Siberia. Significant genetic differences were observed between *T. sibirica* and its most closely related congeneric *T. cordata*. With the morphological differences (Vlasova, 1996; Pigott, 2012) and observed genetic differences, we regard southern Siberian populations as a separate taxon from *T. cordata*, thus supporting our first hypothesis. However, we acknowledge that more work using other markers (chloroplast/nuclear DNA sequencing) is required to fully resolve this issue.

The estimated time of divergence was relatively recent, and when considering a 100 year generation time, it confirms our second hypothesis of a pre-LGM split. It is likely that the split between the two taxa occurred during the Middle Pleistocene in Siberia and that *T. sibirica* has persisted in the region in several refugia.

The molecular markers revealed low genetic diversity within *T. sibirica*, confirming our third hypothesis. Additionally, we observed high levels of clonality within *T. sibirica* and these were greater than clonal levels within *T. cordata*, confirming our fourth hypothesis. Therefore, we strongly recommended that this rare and isolated species is considered an important conservation and management unit, and an effort to restore populations in the area (and within its natural range), should be considered and defined as soon as possible to ensure that these relict populations are preserved into the future.

Acknowledgements:

We would like to thank the anonymous reviewers for their useful comments and suggestions on an earlier version of this manuscript. We are grateful to Jade Lauren Cawthray-Simms for her assistance with sample collection in Poland and Dr Renata Krzyściak-Kosińska for her assistance in gaining permission to collect leaf samples from the Białowieża National Park, Poland. We are also grateful to Drs Jiří Danihelka, Dmitry German, Nikolai Lashchinskyi, Martin Kočí and other colleagues from Masaryk University, Brno, Czech Republic, Altai State University, Barnaul, Russia, and Russian Academy of Sciences, Novosibirsk, Russia, for their help with collecting samples of *Tilia cordata* and *T. sibirica* from Russia. This study was funded by the RB Cooke Studentship at Newcastle University and by a Travelling Fellowship from the Royal Society of Biology (SL).

Author contributions:

SL conceived and designed the study, collected samples, carried out lab work, analysed the data and wrote the paper.

MC and KW collected samples and contributed to writing the paper.

Conflicts of interest: None

References

- Amelin, I.I., Blyakharchuk, T.A., 2016. Rasprostranenie lipy sibirskoi (*Tilia sibirica* Bayer) v Kemerovskoi oblasti [Distribution of Siberian linden (*Tilia sibirica* Bayer) in Kemerovo oblast]. . Vestn. Tomsk. Gos. Univ. Biol. 2, 30-52. [in Russian].
- Arnaud-Haond, S., Belkhir, K., 2007. GenClone: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Mol. Ecol. Notes* 7, 15-17.
- Arnaud-Haond, S., Duarte, C.M., Alberto, F., Serrao, E.A., 2007. Standardizing methods to address clonality in population studies. *Molecular Ecology* 16, 5115-5139.
- Bagnoli, F., Tsuda, Y., Fineschi, S., Bruschi, P., Magri, D., Zhelev, P., Paule, L., Simeone, M.C., González-Martínez, S.C., Vendramin, G.G., 2015. Combining molecular and fossil data to infer demographic history of *Quercus cerris*: insights on European eastern glacial refugia. *Journal of Biogeography*.
- Bhagwat, S.A., Willis, K.J., 2008. Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *Journal of Biogeography* 35, 464-482.
- Blomdin, R., Heyman, J., Stroeven, A.P., Hättestrand, C., Harbor, J.M., Gribenski, N., Jansson, K.N., Petrakov, D.A., Ivanov, M.N., Alexander, O., 2016. Glacial geomorphology of the Altai and Western Sayan Mountains, Central Asia. *Journal of Maps* 12, 123-136.
- Bolikhovskaya, N.S., Shunkov, M.V., 2014. Pleistocene Environments of Northwestern Altai: Vegetation and Climate. *Archaeology, Ethnology and Anthropology of Eurasia* 42, 2-17.
- Bragina, E.V., Radeloff, V.C., Baumann, M., Wendland, K., Kuemmerle, T., Pidgeon, A.M., 2015. Effectiveness of protected areas in the Western Caucasus before and after the transition to post-socialism. *Biological Conservation* 184, 456-464.
- Chapuis, M.-P., Estoup, A., 2007. Microsatellite Null Alleles and Estimation of Population Differentiation. *Molecular Biology and Evolution* 24, 621-631.
- Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.-M., Estoup, A., 2014. DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, btt763.
- Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.-M., Balding, D.J., Guillemaud, T., Estoup, A., 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24, 2713-2719.
- Crawford, N.G., 2010. smogd: software for the measurement of genetic diversity. *Mol. Ecol. Resour.* 10, 556-557.
- Dorken, M.E., Eckert, C.G., 2001. Severely reduced sexual reproduction in northern populations of a clonal plant, *Decodon verticillatus* (Lythraceae). *Journal of Ecology* 89, 339-350.
- Earl, D., vonHoldt, B., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4, 359-361.
- Ellegren, H., Primmer, C.R., Sheldon, B.C., 1995. Microsatellite evolution - Directionality or bias. *Nature Genetics* 11, 360-362.
- Ellstrand, N.C., Elam, D.R., 1993. Population genetic consequences of small population size: implications for plant conservation. *Annual review of Ecology and Systematics*, 217-242.
- Estoup, A., Cornuet, J.M., Dehne-Garcia, A., Loire, E., 2015. New version of the computer program DIYABC (DIYABC v2. 1.0): a user-friendly approach to approximate Bayesian computation for inference on population history using molecular markers.
- Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14, 2611-2620.
- Excoffier, L., Lischer, H.E.L., 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.

- Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567-1587.
- Goldstein, D.B., Ruiz Linares, A., Cavalli-Sforza, L.L., Feldman, M.W., 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139, 463-471.
- Goudet, J., 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *J. Hered.* 86, 485-486.
- Goudet, J., 2001. FSTAT, a program to estimate and test gene diversities and fixation indices version 2.9.3.2, updated from Goudet 1995. FSTAT.
- Groisman, P.Y., Blyakharchuk, T.A., Chernokulsky, A.V., Arzhanov, M.M., Marchesini, L.B., Bogdanova, E.G., Borzenkova, I.I., Bulygina, O.N., Karpenko, A.A., Karpenko, L.V., 2013. Climate changes in Siberia, *Regional Environmental Changes in Siberia and Their Global Consequences*. Springer, pp. 57-109.
- Grubov, V.I., 1940. O sovremennom sostoyanii lipovogo ostrova v predgor'yakh Kuznetskogo Alatau [On the current state of the lime island in the Kuznetskii Alatau foothills]. *Sovetskaya Botanika* 1, 84-85 (in Russian).
- Hais, M., Komprdová, K., Ermakov, N., Chytrý, M., 2015. Modelling the last glacial maximum environments for a refugium of Pleistocene biota in the Russian Altai Mountains, Siberia. *Palaeogeography, Palaeoclimatology, Palaeoecology* 438, 135-145.
- Hamrick, J.L., 2004. Response of forest trees to global environmental changes. *Forest Ecology and Management* 197, 323-335.
- Hedrick, P.W., 2005. A standardized genetic differentiation measure. *Evolution* 59, 1633-1638.
- Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. *Biol. J. Linnean Soc.* 68, 87-112.
- Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801-1806.
- Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I., Andrew, R.L., 2017. The K= 2 conundrum. *Molecular Ecology*.
- Jost, L., 2008. GST and its relatives do not measure differentiation. *Molecular Ecology* 17, 4015-4026.
- Khlonov, Y.P., 1965. Lime and Lime Forests in West Siberia. Nauka Press, Novosibirsk.
- Krapivkina, È.D., 2009. Nemoral'nye relikty vo flore chernevoi taigi Gornoi Shorii [Nemoral Relicts in the Flora of Blackish Taiga in Gornaya Shoriya]. Izdatel'stvo SO RAN Novosibirsk (in Russian).
- Krylov, P.N., 1891. Lipa na predgor'yakh Kuznetskogo Alatau [Lime Tree in the Kuznetskii Alatau Foothills], Tomsk (in Russian).
- Krylov, P.N., 1902. Kratkii ocherk flory Tomskoi gubernii i Altaya [Short Outline of the Tomsk Guberniya and the Altai], Tomsk (in Russian).
- Lashchinskiy, N.N., 2009. Rastitel'nost' Salairskogo kryazha [Vegetation of Salair Mountain Range]. Novosibirsk: Geo. (in Russian).
- Levensen, N.D., Tiffin, P., Olson, M.S., 2012. Pleistocene speciation in the genus *Populus* (Salicaceae). *Systematic biology* 61, 401-412.
- Logan, S.A., Phuekvilai, P., Wolff, K., 2015. Ancient woodlands in the limelight: delineation and genetic structure of ancient woodland species *Tilia cordata* and *Tilia platyphyllos* (Tiliaceae) in the UK. *Tree Genetics & Genomes* 11, 1-12.
- Markova, A.K., Simakova, A.N., Puzachenko, A.Y., 2009. Ecosystems of Eastern Europe at the time of maximum cooling of the Valdai glaciation (24–18kyr BP) inferred from data on plant communities and mammal assemblages. *Quaternary International* 201, 53-59.
- Naidina, O.D., Bauch, H.A., 2001. A Holocene pollen record from the Laptev Sea shelf, northern Yakutia. *Global and Planetary Change* 31, 141-153.
- Nei, M., 1987. *Molecular evolutionary genetics*. Columbia university press.
- Nei, M., Chesser, R.K., 1983. Estimation of fixation indices and gene diversities. *Annals of human genetics* 47, 253-259.

- Normand, S., Treier, U.A., Odgaard, B.V., 2011. commentary: Tree refugia and slow forest development in response to post-LGM warming in North-Eastern European Russia. *Frontiers of Biogeography* 2.
- Novák, J., Trotsiuk, V., Sýkora, O., Svoboda, M., Chytrý, M., 2014. Ecology of *Tilia sibirica* in a continental hemiboreal forest, southern Siberia: An analogue of a glacial refugium of broad-leaved temperate trees? *The Holocene*.
- Parks, J.C., Werth, C.R., 1993. A study of the spatial features of clones in an population of Bracken fern, *Pteridium aquilinum* (Dennstaedtiaceae). *Am. J. Bot.* 80, 537-544.
- Pautasso, M., 2009. Geographical genetics and the conservation of forest trees. *Perspectives in Plant Ecology, Evolution and Systematics* 11, 157-189.
- Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics (Oxford, England)* 28, 2537-2539.
- Phuekvilai, P., 2014. Relicts, refugia, and reticulation: a study of population history, hybrids and phylogeny in the long-lived flowering tree genus *Tilia*. Newcastle University.
- Phuekvilai, P., Wolff, K., 2013. Characterization of microsatellite loci in *Tilia platyphyllos* (Malvaceae) and cross-amplification in related species. *Applications in Plant Sciences* 1, 1200386.
- Pielou, E.C., 1969. *An introduction to mathematical ecology*. Wiley-InterScience, New York.
- Pigott, D., 2012. *Lime-trees and Basswoods*. Cambridge University Press, Cambridge.
- Polozhii, A.V., Krapivkina, E.D., 1985. *Relikty tretichnykh shirokolistvennykh lesov vo flore Sibiri (Relics of the Tertiary deciduous forests in the flora of Siberia)*. Tomsk University Press, Tomsk (in Russian).
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.
- Quantum-G.I.S., 2016. Quantum G.I.S. v2.16.2.
- Raymond, M., Rousset, F., 1995. Genepop (version 1.2) - population genetics software for exact tests and ecumenicism. *J. Hered.* 86, 248-249.
- Reverdatto, L., 1925. K voprosu o sud'be lipovogo ostrova v Kuznetskom Alatau [On the Issue of Fate of the Lime Site in the Kuznetskii Alatau]. Tomsk (in Russian).
- Richardson, J.E., Whitlock, B.A., Meerow, A., Madriñán, S., 2015. The Age of Chocolate: a biogeographic history of Theobroma and Malvaceae. *Frontiers in Ecology and Evolution* 3, 120.
- Rosenberg, N.A., 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.
- Shchur, A., Bragina, E., Sieber, A., Pidgeon, A.M., Radeloff, V.C., 2017. Monitoring selective logging with Landsat satellite imagery reveals that protected forests in Western Siberia experience greater harvest than non-protected forests. *Environmental Conservation* 44, 191-199.
- Simpson, E.H., 1949. Measurement of diversity. *Nature*.
- Stewart, J.R., Lister, A.M., 2001. Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* 16, 608-613.
- Svenning, J.C., Normand, S., Kageyama, M., 2008. Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology* 96, 1117-1127.
- Tarasov, P., Granoszewski, W., Bezrukova, E., Brewer, S., Nita, M., Abzaeva, A., Oberhänsli, H., 2005. Quantitative reconstruction of the last interglacial vegetation and climate based on the pollen record from Lake Baikal, Russia. *Climate Dynamics* 25, 625-637.
- Tzedakis, P.C., Emerson, B.C., Hewitt, G.M., 2013. Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution* 28, 696-704.
- Väliranta, M., Kaakinen, A., Kuhry, P., Kultti, S., Salonen, J.S., Seppä, H., 2011. Scattered late-glacial and early Holocene tree populations as dispersal nuclei for forest development in north-eastern European Russia. *Journal of Biogeography* 38, 922-932.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* 4, 535-538.

Vlasova, N.V., 1996. Semeistvo Tiliaceae – lipovye [Family Tiliaceae], in: Peschkov, G.A. (Ed.), *Flora Sibiri v 14 tomakh: tom 10. Geraniaceae-Cornaceae.*(*Flora Sibiriae: tomus 10 Geraniaceae-Cornaceae*). Nauka Novosibirsk, pp. 65-66 (in Russian).

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.-M., Hoegh-Guldberg, O., Bairlein, F., 2002. Ecological responses to recent climate change. *Nature* 416, 389-395.

Weir, B.S., Cockerham, C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358-1370.

Willis, K.J., Rudner, E., Sümegi, P., 2000. The full-glacial forests of central and southeastern Europe. *Quaternary Research* 53, 203-213.

Willis, K.J., Van Andel, T.H., 2004. Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews* 23, 2369-2387.

Willis, K.J., Whittaker, R.J., 2000. The refugial debate. *Science(Washington)* 287, 1406-1407.

Wolfe, J.A., Wehr, W., 1987. Middle Eocene dicotyledonous plants from Republic, northeastern Washington, - ed. USGS Publications Warehouse.

Table 1 Origin of *Tilia sibirica* and *T. cordata* populations used in this study.

| Species | Country | Population | Code | N | Latitude (°N) | Longitude (°E) |
|--------------------|-------------------------------|-------------------------|------|----|---------------|----------------|
| <i>T. sibirica</i> | Russia, Kemerovo Region | Kuzedeevo ¹ | RK12 | 20 | 53.3210 | 87.2536 |
| | | | RK21 | 15 | 53.3270 | 87.2360 |
| | | | RK22 | 19 | 53.3172 | 87.2938 |
| | | | RK28 | 22 | 53.3168 | 87.2853 |
| | | | RK29 | 21 | 53.3157 | 87.2835 |
| | | | RK38 | 16 | 53.3169 | 87.2482 |
| <i>T. cordata</i> | Austria | Stams ² | AS01 | 23 | 47.2757 | 10.9772 |
| | Poland | Białowieża ³ | PB69 | 17 | 52.7289 | 23.8328 |
| | | | PB99 | 12 | 52.7186 | 23.8443 |
| | | | PB40 | 13 | 52.7336 | 23.8319 |
| | Russia, Tyumen Region | Vagay area ⁴ | RV20 | 20 | 57.5096 | 69.1954 |
| | | | RV25 | 13 | 57.9299 | 68.9097 |

¹ Collected in July 2012 by Milan Chytrý, Jiří Danihelka and Martin Kočí; ² Collected in August 2013 by Kirsten Wolff; ³ Collected in July 2014 by Samuel Logan; ⁴ Collected in August 2013 by Milan Chytrý and Martin Kočí.

Table 2 Historic parameters used for the simple split model (SSM). *N#* – Effective population size; *Na/Na#* – Ancestral population size; *td* – time of divergence.

| Taxa/Parameter | SSM (min – max) |
|--------------------------------|-----------------|
| <i>T. sibirica N#</i> | 10 – 20,000 |
| <i>T. cordata N#</i> | 10 – 20,000 |
| Ancestral population <i>Na</i> | 10 – 20,000 |
| <i>td</i> (in generations) | 10 – 10,000 |

Table 3 Estimates of clonal diversity in six populations of *T. sibirica* (top) and six populations of *T. cordata* from the three regions (bottom) including mean and total (all populations combined). Individuals with missing data were removed from clonal analysis. Number of samples in each population (*N*), number of genotypes (*G*), genotypic richness (*R*), Simpson's index for genotypic diversity (*D*^{*}).

| Pop | <i>N</i> | <i>G</i> | <i>R</i> (<i>G</i> -1/ <i>N</i> -1) | <i>D</i> [*] |
|--------------|--------------|--------------|--------------------------------------|-----------------------|
| RK12 | 20 | 20 | 1.000 | 1.000 |
| RK21 | 15 | 3 | 0.143 | 0.448 |
| RK22 | 19 | 6 | 0.278 | 0.702 |
| RK28 | 22 | 21 | 0.952 | 0.996 |
| RK29 | 21 | 19 | 0.900 | 0.990 |
| RK38 | 16 | 6 | 0.333 | 0.783 |
| Mean | 18.83 | 12.50 | 0.601 | 0.819 |
| Total | 113 | 75 | 0.661 | 0.978 |
| AS01 | 23 | 23 | 1.000 | 1.000 |
| PB69 | 17 | 17 | 1.000 | 1.000 |
| PB99 | 12 | 12 | 1.000 | 1.000 |
| PB40 | 13 | 13 | 1.000 | 1.000 |
| RV20 | 20 | 14 | 0.684 | 0.958 |
| RV25 | 13 | 13 | 1.000 | 1.000 |
| Mean | 16.33 | 15.17 | 0.942 | 0.992 |
| Total | 98 | 92 | 0.938 | 0.998 |

Table 4 Summary statistics of five populations of *T. sibirica*, two Siberian *T. cordata*, three Polish *T. cordata* and one Austrian *T. cordata* population (populations with less than 6 genetic individuals are excluded). N – Number of individuals; N_A – Average number of alleles; A_R – Allelic Richness; H_E – Nei’s unbiased Expected Heterozygosity; H_{E_NULL} – Expected heterozygosity adjusted for null alleles; F_{IS} – Inbreeding coefficient (* $P < 0.05$).

| Population | Species | N | N_A | A_R | H_E | H_{E_NULL} | F_{IS} |
|-------------|--------------------|-------------|-------------|--------------|--------------|---------------|---------------|
| <i>RK12</i> | <i>T. sibirica</i> | 22 | 3.08 | 2.376 | 0.346 | 0.377 | -0.006 |
| <i>RK22</i> | <i>T. sibirica</i> | 6 | 2.00 | 2.000 | 0.274 | 0.266 | 0.095 |
| <i>RK28</i> | <i>T. sibirica</i> | 21 | 2.92 | 2.307 | 0.320 | 0.355 | 0.448* |
| <i>RK29</i> | <i>T. sibirica</i> | 19 | 2.67 | 2.128 | 0.270 | 0.292 | 0.348* |
| <i>RK38</i> | <i>T. sibirica</i> | 6 | 2.25 | 2.250 | 0.379 | 0.368 | -0.194 |
| Mean | | 14.8 | 2.58 | 2.212 | 0.318 | 0.3316 | -0.035 |
| <i>RV20</i> | <i>T. cordata</i> | 14 | 5.08 | 3.878 | 0.597 | 0.577 | 0.158 |
| <i>RV25</i> | <i>T. cordata</i> | 13 | 4.92 | 3.777 | 0.566 | 0.548 | 0.062 |
| <i>PB40</i> | <i>T. cordata</i> | 13 | 5.00 | 3.934 | 0.583 | 0.569 | -0.047 |
| <i>PB69</i> | <i>T. cordata</i> | 17 | 5.83 | 3.835 | 0.523 | 0.510 | -0.033 |
| <i>PB99</i> | <i>T. cordata</i> | 12 | 5.08 | 3.981 | 0.575 | 0.543 | -0.015 |
| <i>AS01</i> | <i>T. cordata</i> | 22 | 5.42 | 3.869 | 0.609 | 0.587 | 0.055 |
| Mean | | 15.2 | 5.22 | 3.879 | 0.576 | 0.556 | 0.030 |

Table 5 Total number of alleles and private alleles within *T. sibirica* and *T. cordata* and per locus genetic differentiation between taxa. A_P - Private alleles within each taxon; F_{ST} (WC) and F_{ST_NULL} (adjusted), D_{est} , $G_{ST\ est}$ and $G'_{ST\ est}$ between taxa.

| Locus | No. of Alleles | A_P <i>T. sibirica</i> | A_P <i>T. cordata</i> | F_{ST} (WC) | F_{ST_NULL} | D_{est} | $G_{ST\ est}$ | $G'_{ST\ est}$ |
|--------------|----------------|--------------------------|-------------------------|---------------|----------------|-----------|---------------|----------------|
| <i>Tc6</i> | 13 | - | 7 | 0.17 | 0.15 | 0.63 | 0.09 | 0.66 |
| <i>Tc937</i> | 8 | 1 | 4 | 0.04 | 0.11 | 0.03 | 0.02 | 0.06 |
| <i>Tc920</i> | 15 | 2 | 10 | 0.39 | 0.29 | 0.85 | 0.25 | 0.88 |
| <i>Tc8</i> | 1 | - | - | - | 0.04 | 0.00 | - | - |
| <i>Tc943</i> | 8 | 2 | 4 | 0.31 | 0.26 | 0.23 | 0.20 | 0.38 |
| <i>Tc4</i> | 10 | - | 6 | 0.17 | 0.16 | 0.38 | 0.10 | 0.43 |
| <i>Tc927</i> | 3 | - | 2 | 0.08 | 0.08 | 0.01 | 0.05 | 0.06 |
| <i>Tc915</i> | 19 | 1 | 12 | 0.15 | 0.17 | 0.52 | 0.08 | 0.56 |
| <i>Tc963</i> | 29 | 3 | 18 | 0.08 | 0.08 | 0.53 | 0.04 | 0.55 |
| <i>Tc5</i> | 19 | 3 | 12 | 0.05 | 0.07 | 0.10 | 0.02 | 0.12 |
| <i>Tc951</i> | 6 | 2 | 2 | 0.14 | 0.09 | 0.13 | 0.08 | 0.20 |
| <i>Tc7</i> | 5 | - | 4 | 0.24 | 0.27 | 0.10 | 0.15 | 0.23 |
| Total/Mean | 136 | 14 | 81 | 0.17 | 0.15 | 0.29 | 0.10 | 0.38 |

Table 6 Population pairwise F_{ST} values and significance between *Tilia sibirica* and *T. cordata* (* - 0.05, ** - 0.01, *** - 0.001, NS - Not significant).

| | sibiri ca | sibiri ca | sibiri ca | sibiri ca | sibiri ca | cord ata | cord ata | cord ata | cord ata | cord ata | cord ata |
|----|--------------|--------------|--------------|--------------|--------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | RK2 | RK2 | RK1 | RK2 | RK3 | RV2 | RV2 | PB6 | PB9 | PB4 | AS0 |
| | 2 | 8 | 2 | 9 | 8 | 0 | 5 | 9 | 9 | 0 | 1 |
| RK | | | | | | | | | | | |
| 22 | - | NS | * | NS | NS | *** | ** | *** | ** | ** | *** |
| RK | 0.01 | | | | | | | | | | |
| 28 | 1 | - | *** | ** | NS | *** | *** | *** | *** | *** | *** |
| RK | 0.07 | 0.10 | | | | | | | | | |
| 12 | 8 | 7 | - | *** | * | *** | *** | *** | *** | *** | *** |
| RK | 0.07 | 0.08 | 0.12 | | | | | | | | |
| 29 | 3 | 8 | 0 | - | ** | *** | *** | *** | *** | *** | *** |
| RK | 0.17 | 0.10 | 0.16 | 0.21 | | | | | | | |
| 38 | 2 | 9 | 8 | 1 | - | ** | ** | *** | ** | ** | *** |
| RV | 0.22 | 0.24 | 0.24 | 0.29 | 0.19 | | | | | | |
| 20 | 2 | 2 | 7 | 1 | 4 | - | NS | *** | *** | *** | *** |
| RV | 0.22 | 0.23 | 0.25 | 0.30 | 0.17 | 0.00 | | | | | |
| 25 | 7 | 4 | 6 | 3 | 5 | 2 | - | *** | *** | *** | *** |
| PB | 0.22 | 0.23 | 0.22 | 0.25 | 0.21 | 0.09 | 0.11 | | | | |
| 69 | 2 | 8 | 7 | 8 | 5 | 4 | 7 | - | ** | ** | *** |
| PB | 0.19 | 0.20 | 0.19 | 0.26 | 0.17 | 0.06 | 0.08 | 0.04 | | | |
| 99 | 5 | 7 | 2 | 0 | 5 | 2 | 3 | 6 | - | NS | *** |
| PB | 0.17 | 0.19 | 0.18 | 0.22 | 0.15 | 0.04 | 0.05 | 0.03 | 0.03 | | |
| 40 | 8 | 0 | 2 | 2 | 1 | 6 | 8 | 1 | 3 | - | *** |
| AS | 0.22 | 0.23 | 0.24 | 0.26 | 0.18 | 0.05 | 0.04 | 0.09 | 0.08 | 0.05 | |
| 01 | 2 | 4 | 3 | 3 | 1 | 0 | 9 | 9 | 5 | 1 | - |

Table 7 Analysis of Molecular Variance (AMOVA) showing the partitioning of genetic variation among taxa, populations and individuals within populations.

| Source of variation | Sum of squares | Variance components | Percentage variation | <i>P</i> value |
|-------------------------------|----------------|---------------------|----------------------|----------------|
| Among taxa | 103.993 | 0.565 | 15.86% | <0.001 |
| Among populations within taxa | 88.622 | 0.243 | 6.83% | <0.001 |
| Within populations | 882.185 | 2.753 | 77.31% | <0.001 |
| Total | 1074.80 | 3.561 | 100% | |

Figure 1 *Tilia cordata* and *T. sibirica* distribution range map with study sites. *T. cordata* (green) range data downloaded from EUFORGEN (www.euforgen.org). *T. sibirica* (orange) range data are based on Amelin and Blyakharchuk (2016). Map constructed in QGIS v2.16.2 (Quantum-G.I.S., 2016).

Figure 2: The simple split model tested in DIYABC: 0 – time of sampling; td – time of divergence; $N1$ – Population 1 (*T. sibirica*); $N2$ – Population 2 (*T. cordata*); N_a – Ancestral population. Time is not to scale.

Figure 3 PCoA of 169 *Tilia* individuals from three regions. Orange diamonds represent *Tilia sibirica* from Kuzedeevo, Russia. Green squares represent *T. cordata* from Vagay, Russia, green triangles represent *T. cordata* from Białowieża, Poland and green circles represent *T. cordata* from Stams, Austria. Axes 1 and 2 together explain 33% of the genetic variation.

Figure 4 Assignment of 169 *Tilia* individuals inferred by Bayesian clustering analysis implemented in STRUCTURE, optimum K averaged across runs using CLUMPP and visualized in DISTRUCT. The orange group represents *T. sibirica* and the green group represents *T. cordata*.

Figure 5 Goodness-of-fit of the SSM, assessed by model check within DIYABC. The PCA shows the observed data set (yellow dot) nested within the posterior predictive distribution (large green dots) and the large cloud of 10,000 (1%) simulated data from the prior.



Fig 1

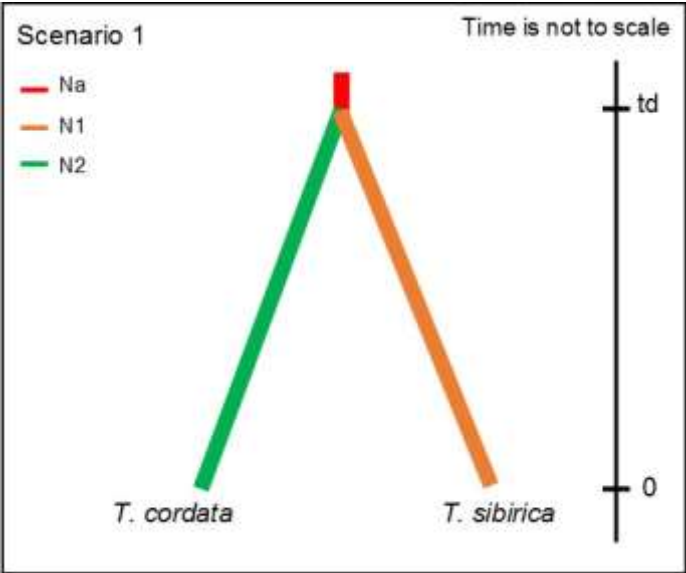


Fig 2

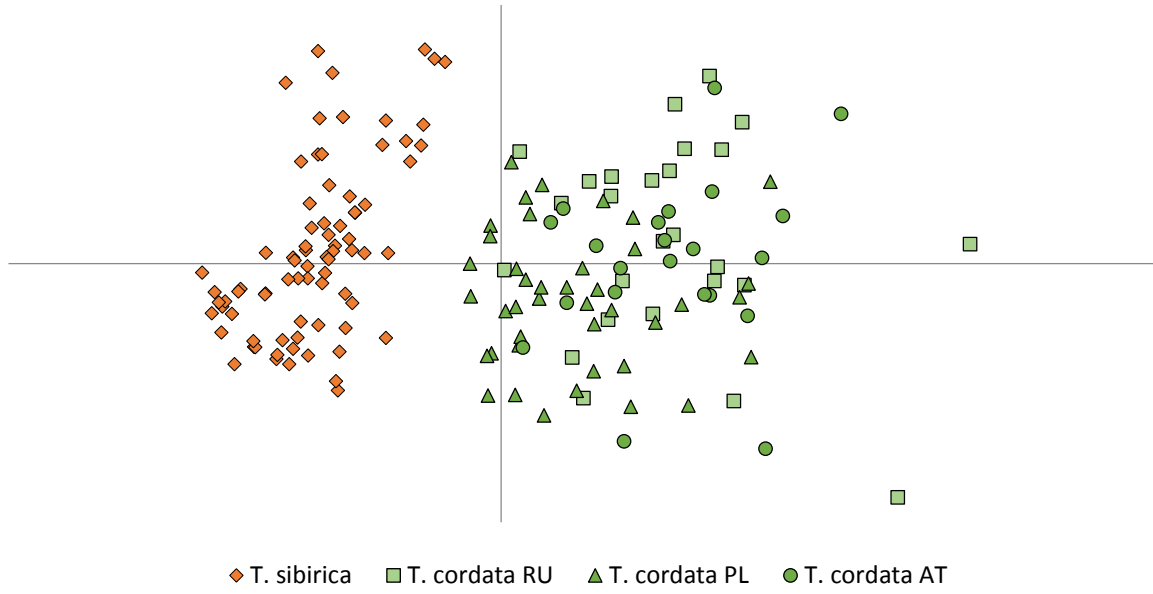


Fig 3

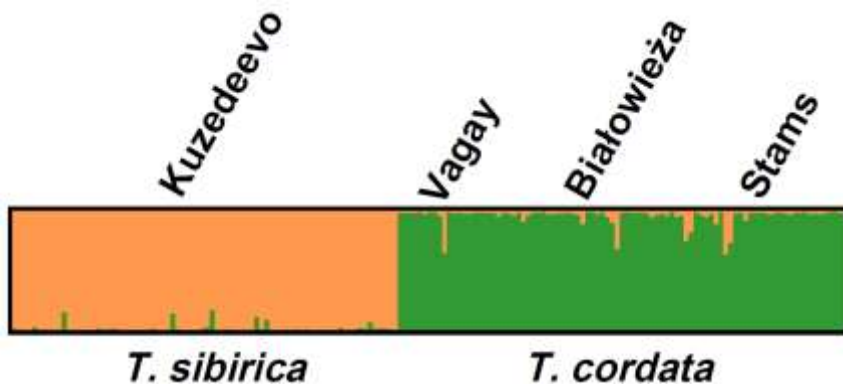


Fig 4

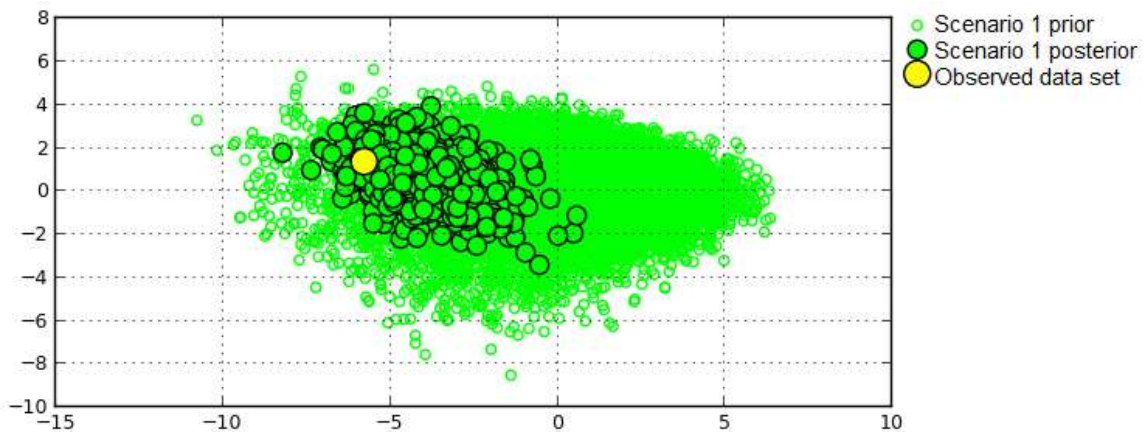
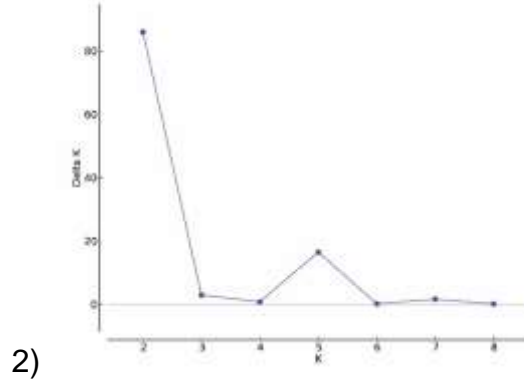
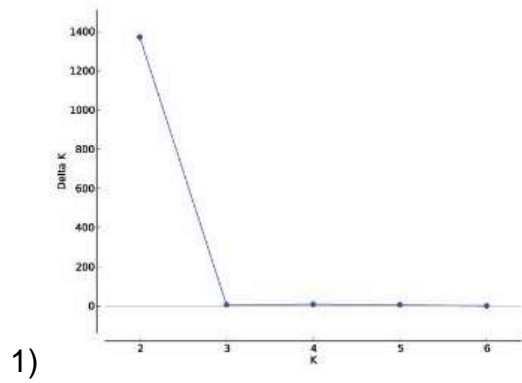


Fig 5

Appendix A $P_{gen} f$ and $P_{sex} f$ values at the first reencounter, for each MLG.

| MLG | $P_{gen} (f)$ | No. of ramets | $P_{sex} (f)$ | MLG | $P_{gen} (f)$ | No. of ramets | $P_{sex} (f)$ |
|-----|---------------|---------------|---------------|-----|---------------|---------------|---------------|
| 1 | -3.27E-08 | | | 38 | 3.17E-05 | | |
| 2 | 6.57E-06 | | | 39 | 9.53E-05 | | |
| 3 | 3.64E-09 | | | 40 | 1.14E-06 | | |
| 4 | 1.09E-06 | | | 41 | 6.21E-05 | | |
| 5 | 3.01E-06 | | | 42 | 4.91E-06 | | |
| 5 | 3.01E-06 | 2 | 0.00034 | 43 | 1.17E-08 | | |
| 6 | 8.01E-07 | | | 44 | 8.10E-06 | | |
| 7 | 7.14E-06 | 3 | 0.000806 | 45 | 1.10E-05 | | |
| 8 | 4.73E-07 | 6 | 5.35E-05 | 45 | 1.10E-05 | 2 | 0.001242 |
| 9 | 9.46E-08 | | | 46 | 1.51E-07 | | |
| 10 | 1.28E-05 | 9 | 0.00145 | 47 | 4.37E-07 | | |
| 11 | 1.34E-07 | | | 48 | 2.42E-06 | | |
| 12 | 4.24E-08 | | | 49 | 3.20E-08 | | |
| 13 | 1.57E-05 | | | 50 | 5.16E-08 | | |
| 14 | 5.50E-06 | | | 51 | 4.82E-07 | | |
| 15 | 2.78E-07 | | | 52 | 4.16E-06 | | |
| 16 | 2.11E-07 | | | 53 | 1.64E-05 | | |
| 17 | 6.79E-07 | | | 54 | 2.46E-07 | | |
| 18 | 4.92E-09 | | | 55 | 0.000133 | | |
| 18 | 4.92E-09 | 2 | 5.56E-07 | 56 | 3.91E-06 | | |
| 19 | 1.34E-07 | 7 | 1.51E-05 | 56 | 3.91E-06 | 2 | 0.000442 |
| 20 | 1.53E-09 | | | 57 | 0.000358 | | |
| 21 | 9.57E-06 | | | 58 | 8.78E-05 | | |
| 22 | 4.29E-06 | | | 59 | 7.62E-05 | | |
| 23 | 2.74E-08 | | | 60 | 0.000852 | | |
| 24 | 1.86E-05 | | | 61 | 1.43E-05 | | |
| 25 | 8.88E-06 | | | 62 | 4.04E-08 | | |
| 26 | 1.62E-06 | | | 63 | 1.27E-05 | | |
| 27 | 6.69E-09 | | | 64 | 7.74E-08 | | |
| 28 | 9.08E-07 | | | 65 | 2.59E-05 | | |
| 29 | 7.64E-07 | | | 66 | 4.10E-08 | | |
| 30 | 3.13E-08 | | | 67 | 2.26E-06 | 3 | 0.000255 |
| 31 | 1.48E-07 | | | 68 | 1.04E-06 | 11 | 0.000118 |
| 31 | 1.48E-07 | 2 | 1.67E-05 | 69 | 3.83E-08 | | |
| 32 | 2.22E-06 | | | 70 | 3.50E-05 | | |
| 33 | 3.83E-08 | | | 71 | 7.55E-08 | | |
| 34 | 1.03E-06 | | | 72 | 2.04E-07 | | |
| 35 | 1.43E-07 | | | 73 | 2.54E-05 | | |
| 36 | 7.54E-05 | | | 74 | 5.35E-05 | | |
| 37 | 1.23E-05 | | | 75 | 4.44E-06 | | |

Appendix B (1) Evanno's ΔK using *T. sibirica* and *T. cordata* showing $K=2$ to be optimal, implemented in STRUCTURE HARVESTER. (2) Evanno's ΔK method of the *T. sibirica* dataset suggesting $K=2$. (3) STRUCTURE output of the assignment of 78 *T. sibirica* individuals showing two genetic groups ($K=2$). (4) STRUCTURE output of the assignment of 78 *T. sibirica* individuals ($K=5$).



Appendix C Estimated null allele frequencies at each locus within each population.

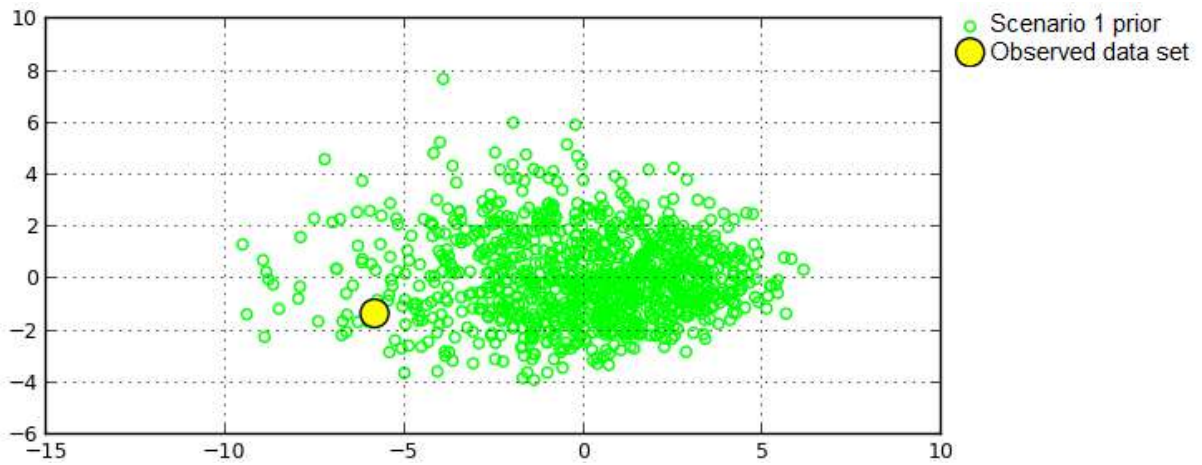
| Estimate of null allele | | | Estimate of null allele | | |
|-------------------------|------|-----------|-------------------------|------|-----------|
| Locus | Pop | frequency | Locus | Pop | frequency |
| Tc6 | RK22 | 0.00000 | Tc8 | RK22 | 0.00100 |
| Tc6 | RK28 | 0.02456 | Tc8 | RK28 | 0.00100 |
| Tc6 | RK12 | 0.00000 | Tc8 | RK12 | 0.00100 |
| Tc6 | RK29 | 0.05909 | Tc8 | RK29 | 0.00100 |
| Tc6 | RK38 | 0.00000 | Tc8 | RK38 | 0.00100 |
| Tc6 | RV20 | 0.00000 | Tc8 | RV20 | 0.00100 |
| Tc6 | RV25 | 0.01465 | Tc8 | RV25 | 0.00100 |
| Tc6 | PB69 | 0.00000 | Tc8 | PB69 | 0.00100 |
| Tc6 | PB99 | 0.08851 | Tc8 | PB99 | 0.00100 |
| Tc6 | PB40 | 0.00000 | Tc8 | PB40 | 0.00100 |
| Tc6 | AS01 | 0.06628 | Tc8 | AS01 | 0.00100 |
| Tc937 | RK22 | 0.00100 | Tc943 | RK22 | 0.00100 |
| Tc937 | RK28 | 0.14980 | Tc943 | RK28 | 0.00001 |
| Tc937 | RK12 | 0.00000 | Tc943 | RK12 | 0.00100 |
| Tc937 | RK29 | 0.00100 | Tc943 | RK29 | 0.00100 |
| Tc937 | RK38 | 0.00000 | Tc943 | RK38 | 0.00000 |
| Tc937 | RV20 | 0.08611 | Tc943 | RV20 | 0.04159 |
| Tc937 | RV25 | 0.00002 | Tc943 | RV25 | 0.00001 |
| Tc937 | PB69 | 0.00000 | Tc943 | PB69 | 0.00005 |
| Tc937 | PB99 | 0.00000 | Tc943 | PB99 | 0.00000 |
| Tc937 | PB40 | 0.00000 | Tc943 | PB40 | 0.09021 |
| Tc937 | AS01 | 0.00000 | Tc943 | AS01 | 0.11803 |
| Tc920 | RK22 | 0.00100 | Tc4 | RK22 | 0.33333 |
| Tc920 | RK28 | 0.06006 | Tc4 | RK28 | 0.31478 |
| Tc920 | RK12 | 0.07644 | Tc4 | RK12 | 0.14131 |
| Tc920 | RK29 | 0.12766 | Tc4 | RK29 | 0.27846 |
| Tc920 | RK38 | 0.00001 | Tc4 | RK38 | 0.25122 |
| Tc920 | RV20 | 0.01141 | Tc4 | RV20 | 0.13654 |
| Tc920 | RV25 | 0.00000 | Tc4 | RV25 | 0.00000 |
| Tc920 | PB69 | 0.00000 | Tc4 | PB69 | 0.02914 |
| Tc920 | PB99 | 0.00000 | Tc4 | PB99 | 0.00000 |
| Tc920 | PB40 | 0.00000 | Tc4 | PB40 | 0.08336 |
| Tc920 | AS01 | 0.00000 | Tc4 | AS01 | 0.00001 |

Appendix C (cont.) Estimated null allele frequencies at each locus within each population

| Locus | Pop | Estimate of null allele frequency | Locus | Pop | Estimate of null allele frequency |
|-------|------|-----------------------------------|-------|------|-----------------------------------|
| Tc927 | RK22 | 0.00100 | Tc5 | RK22 | 0.00000 |
| Tc927 | RK28 | 0.00100 | Tc5 | RK28 | 0.34704 |
| Tc927 | RK12 | 0.00100 | Tc5 | RK12 | 0.22950 |
| Tc927 | RK29 | 0.00100 | Tc5 | RK29 | 0.15511 |
| Tc927 | RK38 | 0.00100 | Tc5 | RK38 | 0.00000 |
| Tc927 | RV20 | 0.00000 | Tc5 | RV20 | 0.15313 |
| Tc927 | RV25 | 0.00002 | Tc5 | RV25 | 0.03213 |
| Tc927 | PB69 | 0.00001 | Tc5 | PB69 | 0.00000 |
| Tc927 | PB99 | 0.00000 | Tc5 | PB99 | 0.00000 |
| Tc927 | PB40 | 0.00002 | Tc5 | PB40 | 0.00000 |
| Tc927 | AS01 | 0.00001 | Tc5 | AS01 | 0.00000 |
| Tc915 | RK22 | 0.00000 | Tc951 | RK22 | 0.00000 |
| Tc915 | RK28 | 0.14980 | Tc951 | RK28 | 0.00001 |
| Tc915 | RK12 | 0.00000 | Tc951 | RK12 | 0.09875 |
| Tc915 | RK29 | 0.03529 | Tc951 | RK29 | 0.00001 |
| Tc915 | RK38 | 0.05136 | Tc951 | RK38 | 0.00000 |
| Tc915 | RV20 | 0.02828 | Tc951 | RV20 | 0.04728 |
| Tc915 | RV25 | 0.05138 | Tc951 | RV25 | 0.00004 |
| Tc915 | PB69 | 0.00000 | Tc951 | PB69 | 0.00000 |
| Tc915 | PB99 | 0.00000 | Tc951 | PB99 | 0.00000 |
| Tc915 | PB40 | 0.00000 | Tc951 | PB40 | 0.00000 |
| Tc915 | AS01 | 0.03224 | Tc951 | AS01 | 0.01048 |
| Tc963 | RK22 | 0.00000 | Tc7 | RK22 | 0.00100 |
| Tc963 | RK28 | 0.21021 | Tc7 | RK28 | 0.00100 |
| Tc963 | RK12 | 0.00000 | Tc7 | RK12 | 0.00100 |
| Tc963 | RK29 | 0.13892 | Tc7 | RK29 | 0.00100 |
| Tc963 | RK38 | 0.00000 | Tc7 | RK38 | 0.00100 |
| Tc963 | RV20 | 0.01352 | Tc7 | RV20 | 0.14709 |
| Tc963 | RV25 | 0.00000 | Tc7 | RV25 | 0.00836 |
| Tc963 | PB69 | 0.00000 | Tc7 | PB69 | 0.00003 |
| Tc963 | PB99 | 0.19397 | Tc7 | PB99 | 0.00000 |
| Tc963 | PB40 | 0.00000 | Tc7 | PB40 | 0.00000 |
| Tc963 | AS01 | 0.04006 | Tc7 | AS01 | 0.00001 |

Appendix D F_{ST_NULL} - Population pairwise F_{ST} values between the two taxa adjusted for null alleles

| | sibirica | sibirica | sibirica | sibirica | sibirica | cordata | cordata | cordata | cordata | cordata | cordata |
|------|----------|----------|----------|----------|----------|---------|---------|---------|---------|---------|---------|
| | RK22 | RK28 | RK12 | RK29 | RK38 | RV20 | RV25 | PB69 | PB99 | PB40 | AS01 |
| RK22 | - | | | | | | | | | | |
| RK28 | 0.042 | - | | | | | | | | | |
| RK12 | 0.088 | 0.093 | - | | | | | | | | |
| RK29 | 0.078 | 0.065 | 0.109 | - | | | | | | | |
| RK38 | 0.174 | 0.112 | 0.158 | 0.210 | - | | | | | | |
| RV20 | 0.232 | 0.231 | 0.240 | 0.279 | 0.197 | - | | | | | |
| RV25 | 0.231 | 0.223 | 0.244 | 0.286 | 0.175 | 0.005 | - | | | | |
| PB69 | 0.214 | 0.214 | 0.209 | 0.231 | 0.206 | 0.100 | 0.115 | - | | | |
| PB99 | 0.196 | 0.190 | 0.181 | 0.238 | 0.174 | 0.067 | 0.086 | 0.046 | - | | |
| PB40 | 0.173 | 0.166 | 0.167 | 0.200 | 0.143 | 0.052 | 0.059 | 0.031 | 0.037 | - | |
| AS01 | 0.220 | 0.214 | 0.230 | 0.245 | 0.176 | 0.052 | 0.050 | 0.099 | 0.085 | 0.051 | - |



Appendix E SSM pre-evaluate scenario and priors using all available summary statistics. The PCA graph shows the observed dataset (yellow dot) positioned within 1000 simulated datasets (green dots) suggesting the statistics are suitable.

Appendix F SSM using all available summary statistics: The proportion of simulated data sets that have values below the observed data set are presented. One-sample statistics – mean number of alleles (A), genic diversity (H), allele size variance (V), Garza-Williamson M (MGW); Two-sample statistics - mean number of alleles ($A2P$), genic diversity ($H2P$), allele size variance ($V2P$), genetic differentiation (F_{ST}), the classification index (LIK), shared allele distance (DAS) and genetic distance between two samples $(\delta\mu)^2$. (*) = 0.05.

| Summary statistic | Observed value | P (sim<obs) |
|--------------------------|----------------|---------------|
| A_{1_1} | 4.6667 | 0.0940 |
| A_{1_2} | 6.0833 | 0.2008 |
| H_{1_1} | 0.3541 | 0.0470 (*) |
| H_{1_2} | 0.5841 | 0.1449 |
| V_{1_1} | 7.0080 | 0.1264 |
| V_{1_2} | 7.0208 | 0.1271 |
| MGW_{1_1} | 0.6054 | 0.2597 |
| MGW_{1_2} | 0.6606 | 0.6757 |
| $A2P_{1_1\&2}$ | 8.1667 | 0.0891 |
| $H2P_{1_1\&2}$ | 0.4621 | 0.0240 (*) |
| $V2P_{1_1\&2}$ | 7.6963 | 0.0645 |
| $F_{ST_{1_1\&2}}$ | 0.2355 | 0.7830 |
| $LIK_{1_1\&2}$ | 1.3834 | 0.0178 (*) |
| $LIK_{1_2\&1}$ | 2.3229 | 0.1055 |
| $DAS_{1_1\&2}$ | 0.4045 | 0.9851 (*) |
| $(\delta\mu)^2_{1_1\&2}$ | 3.7333 | 0.1068 |

Appendix G Parameter estimates of the simple split model: $N1$ – effective population size of *T. sibirica*; $N2$ - effective population size of *T. cordata*; N_a – effective population size of an ancestral population; td – time of divergence in generations; $\hat{A}\mu_{mic}$ – mean SSR mutation rate; $pmic$ – Mean P (parameter of geometric distribution); $snimic$ – mean single nucleotide insertion (SNI).

| SSM model | | | | Quartile | | | | | |
|--------------------|----------|----------|--------------|----------|-------------|----------|----------|--------------|----------|
| Parameter | mean | median | mode | 2.5% | 5% | 25% | 75% | 95% | 97.5% |
| $N1$ | 5720 | 5380 | 4760 | 1330 | 1770 | 3640 | 7370 | 10900 | 12300 |
| $N2$ | 13000 | 13500 | 13700 | 5210 | 6480 | 10800 | 15700 | 18000 | 18500 |
| td | 5150 | 5070 | 4470 | 1110 | 1540 | 3430 | 6850 | 9010 | 9450 |
| N_a | 14200 | 14900 | 18200 | 5420 | 6720 | 11600 | 17500 | 19500 | 19800 |
| $\hat{A}\mu_{mic}$ | 6.46E-05 | 5.01E-05 | 3.86E-05 | 1.46E-05 | 1.78E-05 | 3.31E-05 | 7.79E-05 | 1.57E-04 | 1.99E-04 |
| $pmic$ | 0.637 | 0.632 | 0.585 | 0.291 | 0.335 | 0.499 | 0.777 | 0.948 | 0.981 |
| $snimic$ | 2.18E-06 | 9.51E-07 | 1.57E-08 | 1.85E-08 | 2.74E-08 | 1.96E-07 | 3.40E-06 | 8.15E-06 | 9.16E-06 |

Appendix H Model check (ranked approach) on 1,000 simulated data from the SSM. One sample statistics - mean number of alleles (A), genic diversity (H), allele size variance (V), Garza-Williamson M (MGW); Two-sample statistics - mean number of alleles ($A2P$), genic diversity ($H2P$), allele size variance ($V2P$), genetic differentiation (F_{ST}), the classification index (LIK), shared allele distance (DAS) and genetic distance between two samples $(\delta\mu)^2$. (*) = 0.05.

| Summary statistic | Observed value | P (sim<obs) |
|---------------------------|----------------|---------------|
| A_{1_1} | 4.6667 | 0.3315 |
| A_{1_2} | 6.0833 | 0.4915 |
| H_{1_1} | 0.3541 | 0.0190 (*) |
| H_{1_2} | 0.5841 | 0.2005 |
| V_{1_1} | 7.0080 | 0.3000 |
| V_{1_2} | 7.0208 | 0.1850 |
| MGW_{1_1} | 0.6054 | 0.7285 |
| MGW_{1_2} | 0.6606 | 0.7940 |
| $A2P_{1_1\&2}$ | 8.1667 | 0.4985 |
| $H2P_{1_1\&2}$ | 0.4621 | 0.0175 (*) |
| $V2P_{1_1\&2}$ | 7.6963 | 0.2140 |
| $F_{ST_{1_1\&2}}$ | 0.2355 | 0.7490 |
| $LIK_{1_1\&2}$ | 1.3834 | 0.0780 |
| $LIK_{1_2\&1}$ | 2.3229 | 0.3670 |
| $DAS_{1_1\&2}$ | 0.4045 | 0.9570 (*) |
| $(\delta\mu)^2_{1_1\&2}$ | 3.7333 | 0.2940 |

