

275 years of forestry meets genomics in *Pinus sylvestris*

Tanja Pyhäjärvi^{1,2}  | Sonja T Kujala³  | Outi Savolainen^{1,2} 

¹Department of Ecology and Genetics, University of Oulu, Oulu, Finland

²Biocenter Oulu, University of Oulu, Oulu, Finland

³Natural Resources Institute Finland, Luke, Oulu, Finland

Correspondence

Tanja Pyhäjärvi, Department of Ecology and Genetics, University of Oulu, P.O. Box 3000, 90014 Oulu, Finland.

Email: tanja.pyhajarvi@oulu.fi

Funding information

Biotieteiden ja Ympäristön Tutkimuksen Toimikunta, Grant/Award Number: 287431, 293819, 307582 and 309978

Abstract

Pinus sylvestris has a long history of basic and applied research that is relevant for both forestry and evolutionary studies. Its patterns of adaptive variation and role in forest economic and ecological systems have been studied extensively for nearly 275 years, detailed demography for a 100 years and mating system more than 50 years. However, its reference genome sequence is not yet available and genomic studies have been lagging compared to, for example, *Pinus taeda* and *Picea abies*, two other economically important conifers. Despite the lack of reference genome, many modern genomic methods are applicable for a more detailed look at its biological characteristics. For example, RNA-seq has revealed a complex transcriptional landscape and targeted DNA sequencing displays an excess of rare variants and geographically homogeneously distributed molecular genetic diversity. Current DNA and RNA resources can be used as a reference for gene expression studies, SNP discovery, and further targeted sequencing. In the future, specific consequences of the large genome size, such as functional effects of regulatory open chromatin regions and transposable elements, should be investigated more carefully. For forest breeding and long-term management purposes, genomic data can help in assessing the genetic basis of inbreeding depression and the application of genomic tools for genomic prediction and relatedness estimates. Given the challenges of breeding (long generation time, no easy vegetative propagation) and the economic importance, application of genomic tools has a potential to have a considerable impact. Here, we explore how genomic characteristics of *P. sylvestris*, such as rare alleles and the low extent of linkage disequilibrium, impact the applicability and power of the tools.

KEYWORDS

adaptation, allele frequency spectrum, breeding, genomic prediction, genomics, inbreeding depression, linkage disequilibrium, *Pinus sylvestris*

1 | INTRODUCTION

Pinus sylvestris (Scots pine) is one of the world's most widely distributed and northern conifers reaching from the British Isles in the west to the Siberian taiga in the east. It is also found in the mountainous areas of Mediterranean peninsulas in Southern

Europe. It is especially competitive in poor soils and in dry and extremely cold environments, and compared to many other *Pinus*, its ecological niche is wide (Rehfeldt et al., 2002; Svenning, Normand, & Kageyama, 2008). Especially in the northern part of its distribution, it is a dominant tree in forested areas (Durrant, Rigo, & Caudullo, 2016) and has an important role as a part of forest

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

ecosystems as a global carbon reservoir and also via interactions with soil microbes and fungi (Högberg et al., 2001; Lindén et al., 2014; Pan et al., 2011). Economically, it is an important tree species in Northern Eurasia as raw material for paper and pulp industry and as timber (Durrant et al., 2016; Gardner, 2013; Mason & Alía, 2000; Mullin et al., 2011). It is widely planted for timber outside its natural range. *P. sylvestris* is estimated to cover over 145 million hectares of forest in Eurasia (Durrant et al., 2016; Mason & Alía, 2000; Mullin et al., 2011).

Very rough estimates of the actual population census size can be made by combining distribution and density information. In commercial *P. sylvestris* forests, there are about 2000 stems per hectare after precommercial thinning (Fahlvik, Ekö, & Pettersson, 2005; Väisänen, Kellomäki, Oker-Blom, & Valtonen, 1989), which yields an estimate of population census size of 290×10^9 trees. *P. sylvestris* is not globally under threat in terms of species viability as such (Gardner, 2013). However, its dominant role in large forest ecosystems means that any changes in its distribution or mortality are likely to have large ecological and economic consequences.

Multiple subspecies and varieties of *P. sylvestris* have been described based on morphological and phenological differences (Molotkov & Patlaj, 1991; Pravdin, 1969; Ruby & Wright, 1976), but groupings are not consistent across studies and vary depending on the studied traits (Shutyaev & Giertych, 2000). Often the underlying phenotypic differences have more clinal rather than discrete distribution and the description of separate varieties reflects more the sampling design and silvicultural purposes than true biological groups (Langlet, 1971; Shutyaev & Giertych, 2000). At the genome-wide level, molecular genetic differentiation among sampling locations at nuclear loci is low ($F_{ST} = 0.02$) in most parts of the distribution (Karhu et al., 1996; Kujala & Savolainen, 2012; Muona & Harju, 1989; Prus-Głowacki, Urbaniak, Bujas, & Curtu, 2012; Pyhäjärvi et al., 2007; Tyrmi et al., 2019).

Due to its dominance and economical importance, phenotypic variation of *P. sylvestris* has been of interest since the early days of biological sciences. The first studies of phenotypic and adaptive differences among populations originating from different geographic regions were motivated by shipbuilding and forestry (Alberto et al., 2013; Engler, 1913; Langlet, 1971; Morgenstern, 1996). The first evidence of *P. sylvestris* provenance (common garden) trials in France comes from as early as 1745 (Langlet, 1971). Later, several variably documented national (Eiche, 1966; Heikinheimo, 1950) and international (Giertych & Oleksyn, 1992; Shutyaev & Giertych, 1998; Wiedemann, 1930) provenance trials have been established mainly for silvicultural purposes, also outside its natural range (Wright & Ira, 1962). Even though not originally designed for the purpose, these data have proved valuable for evaluating the level of ecological adaptation (Savolainen, Pyhäjärvi, & Knürr, 2007) and making predictions on responses to climate change (Persson, 1998; Rehfeldt et al., 2002). In the 20th century, very detailed studies on reproductive biology (Koski, 1970; Lönnroth, 1926; Sarvas, 1962) formed a solid basis for modern applied and evolutionary genetic research of *P. sylvestris*.

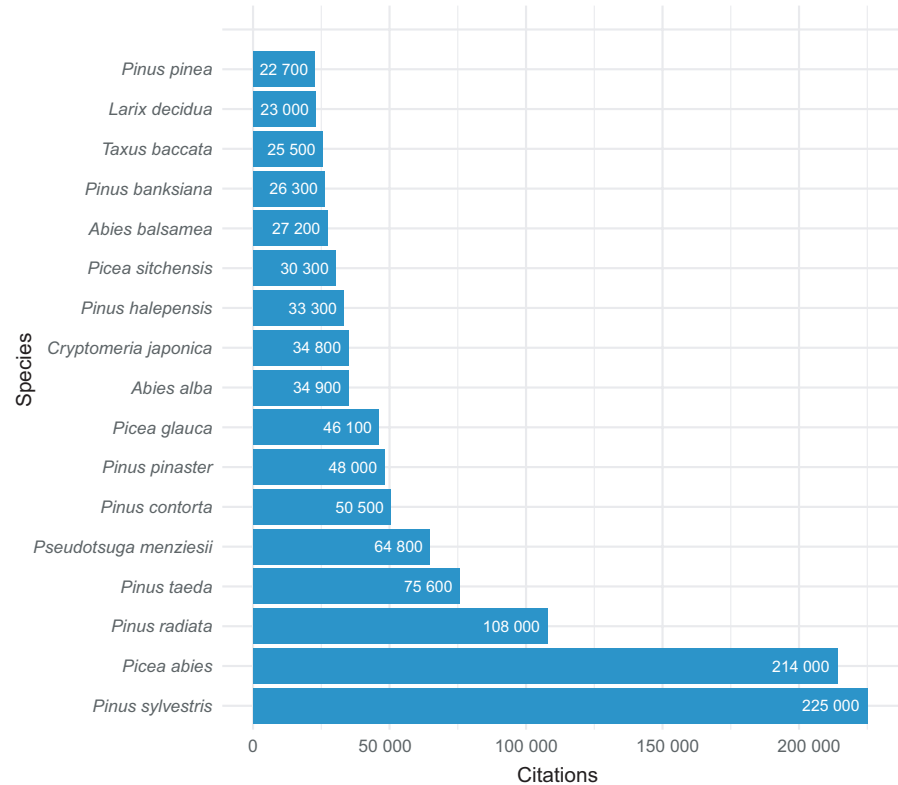
The extensive distribution, large effective population size (N_e), efficient gene flow, predominantly outcrossing nature, and large 22 Gbp genome, yet strong phenotypic differentiation in adaptive traits, make *P. sylvestris* an intriguing system to examine adaptive processes of polygenic traits. Lack of significant longitudinal structure allows replicating sampling along latitudinal transects and inspecting whether the same loci or variants are participating in the adaptive trait variation throughout the distribution or whether the adaptation has emerged via different combinations of adaptive alleles. In comparison with, for example, the North American interior spruce complex (Yeaman et al., 2016), *P. sylvestris* does not significantly hybridize with other species (see, however, Wachowiak & Prus-Głowacki, 2008), likely because most close relatives have small distributions or do not overlap with *P. sylvestris*. Lack of hybridization should further facilitate understanding of the genetic basis of local adaptation. The species is also known to harbor large number of lethal equivalents (Kärkkäinen, Koski, & Savolainen, 1996; Koski, 1971), which allows examining the joint effects of deleterious variation with adaptive and phenotypic variations.

In comparison with farm animals, crops, and vegetables, many coniferous forest trees gene pools have not been extensively changed by humans, natural regeneration is common in large part of the distribution, and breeding is conducted in a handful of countries (Mullin et al., 2011). In Finland and Sweden, breeding efforts on *P. sylvestris* were started in the late 1940s by collecting phenotypically superior plus trees from natural and cultivated forests. Plus trees were used to establish seed orchards for seed production. Later, the genetic quality of the seed orchards has been improved by selection among plus trees based on progeny testing, forming the “1.5 generation” seed orchards (Haapanen, Hynynen, Ruotsalainen, Siipilehto, & Kilpeläinen, 2016; Jansson, Hansen, Haapanen, Kvaalen, & Steffenrem, 2017). With traditional breeding methods, genetic gains up to ~25% for growth traits like volume, and reduction in rotation times have been obtained and resulted in significant economic gains in Finland and Scandinavia (Ahtikoski, Ojansuu, Haapanen, Hynynen, & Kärkkäinen, 2012; Jansson et al., 2017).

Given the low number of breeding cycles conducted so far, there should be plenty of functional genetic variation, and thus, large genetic gains can be expected. By adding genomic methods to the available tools, we can expect further economic gains per unit of time. Genomic information can improve the breeding efforts in *P. sylvestris* through genomewide association analyses (GWAS), inference of genetic relatedness, and by genomic prediction (Isik, 2014; Isik et al., 2016; Meuwissen, Hayes, & Goddard, 2001). So far, efforts to make *P. sylvestris* to flower at earlier age have not been successful. Traits that can be measured only on adult trees hold greatest potential in terms of shortening the breeding cycle (Isik, 2014). Genomic prediction can deal with the highly polygenic basis of trait variation (with unknown underlying loci) unlike other breeding methods, such as gene editing, requiring detailed molecular information.

Given its economical (Praciak, 2013) and ecological importance, and the long record of research and knowledge (Giertych & Matyas, 1991), it is unfortunate that a whole genome sequence

FIGURE 1 Number of approximate Google Scholar citations (as of October 19th 2018) for some most intensively studied conifer species. Number of results indicated in each bar



assembly of *P. sylvestris* has not yet been presented. Even though in many other ways, it is one of the most thoroughly studied conifers (Figure 1); from genomics point of view, it is still a nonmodel organism. However, whole genome sequencing and reference genomes are not a prerequisite for using current methodologies for genomic analysis of *P. sylvestris*.

In this review, we consider biological, life history, and genomic characteristics that are relevant for future applications of genomic information in *P. sylvestris*. We will address (a) mating system and inbreeding depression, (b) genetic diversity and linkage disequilibrium, (c) population structure, history, and allele frequency distribution, (d) adaptive variation at trait and molecular level, (e) genome size and architecture, and (f) application of genomics in breeding and conservation.

2 | MATING SYSTEM AND INBREEDING DEPRESSION

Pinus sylvestris, as many other forest trees, is wind-pollinated, diploid, and predominantly outcrossing. The selfing rate measured in mature seeds is about 5%–10% (Muona, Yazdani, & Rudin, 1987; Rudin, Muona, & Yazdani, 1986). *P. sylvestris* is also known to display high inbreeding depression, also a common observation in forest trees. Inbreeding depression is normally measured by comparing the performance of selfed (or other inbred) offspring to that of controlled outcrossed or open-pollinated offspring. In pines, embryonic mortality can be evaluated by examining the proportion of empty seeds

from different types of crosses (Kärkkäinen et al., 1996; Koski, 1971). The average mortality in selfed seeds is 75%–85% (Kärkkäinen et al., 1996; Koelewijn, Koski, & Savolainen, 1999; Koski, 1971), compared to 20%–30% for open-pollinated seed. Based on these kinds of data, with simplifying assumptions, the number of embryonic lethal equivalents in *P. sylvestris* has been estimated to be 9.4 (Koski, 1971). The level is much higher than in most animals and angiosperms and similar to many other conifers, such as *Picea abies*, *Picea glauca*, or *Pseudotsuga menziesii*, while some pines with limited distributions have much lower levels (e.g., *Pinus radiata* and *Pinus resinosa*; Lynch & Walsh, 1998; Williams & Savolainen, 1996). The selfing rate at fertilization is thus much higher than at the mature seed stage. Further, the selfing rate at fertilization can be even higher because of polyzygotic (simple) polyembryony: An ovule may contain up to four embryos, which can be of selfed or outcrossed origin. These embryos share their maternal haplotype but have been fertilized by different pollen. Only one dominant embryo will survive (Hedrick & Muona, 1990; Koski, 1970; Sorensen, 1982). Inbred mortality during early embryogenesis and replacement by outcrossed progeny results in reduced loss of maternal resources, thus reducing post-seed-maturation inbreeding depression (Kärkkäinen & Savolainen, 1993; Sarvas, 1962).

Other studies (Koelewijn et al., 1999; Muona et al., 1987; Yazdani, Muona, Rudin, & Szmids, 1985) have shown that the low fitness of selfs continue during later years, such that already in a few years old seedlings and especially in young adults, the selfed progeny has been mostly eliminated and the genotypes are in Hardy-Weinberg equilibrium. The results on conifers in general suggest that the overall inbreeding depression is due to a large

number of deleterious, partially recessive alleles (Williams & Savolainen, 1996). Lande, Schemske, and Schultz (1994) showed that the lethals can be maintained in the population despite the selfing when there is a high per genome per generation mutation rate to the deleterious alleles. In addition, stabilizing selection on quantitative traits can interfere in complex ways with selection on lethals and either increase or decrease the probability of purging (Lande & Porcher, 2017).

Genomics can shed light on the genetic architecture of inbreeding depression by characterizing the numbers and effect sizes of deleterious alleles, for example, by mapping (Hedrick & Muona, 1990; Ritland, 1996), and by analyzing heterozygosity of the selfed progeny, as in *Eucalyptus grandis* (Hedrick, Hellsten, & Grattapaglia, 2016): In the selfed progeny, heterozygosity was much higher than the neutral expectation, suggesting that overall selection against homozygotes in selfed offspring was very high ($s = 0.47$). Furthermore, inbreeding depression likely was due to partially deleterious alleles at more than 100 loci, even if overdominance effects could not be fully excluded. Similar methods could help to identify genomic areas with lethals also in conifers. So far, various functional and evolutionary prediction models of allelic substitution have been used to identify deleterious alleles. This has then allowed comparisons of allelic and genotypic frequencies between populations or species, as in *Populus* (Zhang, Zhou, Bawa, Suren, & Holliday, 2016) or in *Picea* (Conte et al., 2017). Estimation of the distribution of fitness effects also allows conclusions on the nature of deleterious alleles. Allele frequency spectrum (AFS) of some 400 loci suggested that *P. sylvestris* has fewer slightly deleterious alleles and a larger proportion of more highly deleterious alleles than other conifers and plants in

general (Grivet et al., 2017; Hodgins, Yeaman, Nurkowski, Rieseberg, & Aitken, 2016).

While mating between relatives can be an important breeding tool for some species, in conifers deleterious alleles are so numerous that breeding strategies using inbreeding will often lead to fixation of deleterious alleles, as shown in simulations by Wu, Hallingbäck, and Sánchez (2016). However, in some cases, variation in levels of inbreeding depression may allow using mating between relatives (Ford, McKeand, Jett, & Isik, 2014). Because of the mostly harmful effects of mating between relatives, it is important to manage inbreeding levels in breeding programs by measuring relatedness with genetic tools.

3 | GENETIC DIVERSITY AND LINKAGE DISEQUILIBRIUM

In many aspects, patterns of genetic diversity in *P. sylvestris* match the expectations for a wind-pollinated tree with large population size. Population structure is almost nonexistent, linkage disequilibrium (LD), the nonrandom association of alleles generally does not extend far, and genotypes are in Hardy–Weinberg equilibrium as expected under random mating (Kujala & Savolainen, 2012; Muona & Szmidi, 1985; Pyhäjärvi et al., 2007; Tyrmi et al., 2019; Wachowiak, Salmela, Ennos, Iason, & Cavers, 2011).

Comparative studies have indicated that *Pinus* in general do not have particularly low nucleotide diversity compared to other plants (Chen, Glémin, & Lascoux, 2017; Eckert et al., 2013; Figure 2). Silent site diversity in the *P. sylvestris* genic regions (based on Sanger

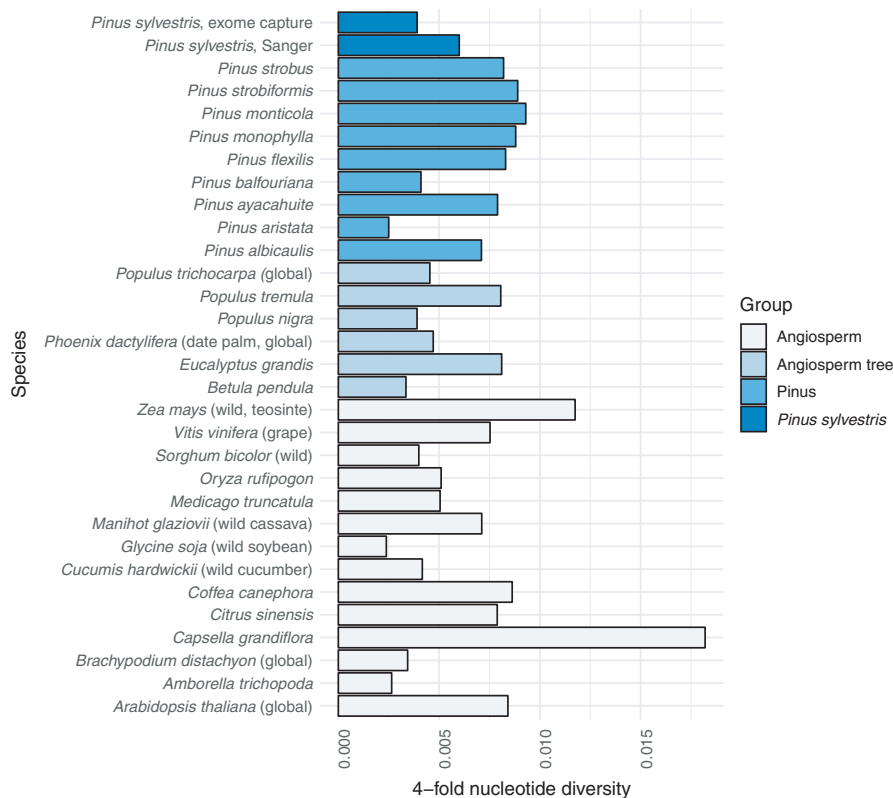


FIGURE 2 4-fold site nucleotide diversity for variety of angiosperms and *Pinus* (data from Chen et al., 2017 except for *Pinus sylvestris*)

sequencing) seems to converge to 0.006/bp (Dvornyk, Sirviö, Mikkonen, & Savolainen, 2002; García-Gil, Mikkonen, & Savolainen, 2003; Grivet et al., 2017; Kujala & Savolainen, 2012; Pyhäjärvi et al., 2007; Wachowiak et al., 2011), and the first genomewide estimate of silent, fourfold degenerate sites genetic diversity is 0.004/bp (Tyrmi et al. 2019). This estimate is slightly lower than the variation observed in other *Pinus* at fourfold degenerate sites (Eckert et al., 2013) and also slightly lower than diversity observed in silent sites of *P. sylvestris* in earlier studies. Note that even silent and fourfold degenerate nucleotide diversity and mutation rate estimates for *P. sylvestris* are based on data derived from genic regions. The overall patterns of diversity further away from genes may be very different and likely higher.

As for many other species, there is a mismatch between the observed nucleotide diversity and the census size in *P. sylvestris* (Lewontin, 1974). Under the assumption of standard neutral equilibrium, $4N_e\mu$ (where N_e is the effective population size and μ is mutation rate per site per generation) is expected to equal the pairwise nucleotide diversity, θ (Tajima, 1983). N_e can then be estimated with $N_e = \theta/(4\mu)$. The mutation rates per bp per year vary from 0.22 to 1.3×10^{-9} (Buschiazzo, Ritland, Bohlmann, & Ritland, 2012; De La Torre, Li, Peer, & Ingvarsson, 2017; Pyhäjärvi et al., 2007; Willyard, Syring, Gernandt, Liston, & Cronn, 2007). Together with silent nucleotide diversity of 0.004/bp and assuming generation time of 20 years, this yields population size estimates ranging from 38,000 to 230,000 individuals. Yet, the actual census size is in the scale of billions of individuals in a seemingly random mating population. Clearly, some of the above-mentioned assumptions are violated. Potential reasons for the discrepancy are as follows: (a) uneven fecundity among individual trees, which would increase the allele frequency variance across the generations, leading to increased drift, and deviations from the assumed Wright–Fisher population model; (b) much lower mutation rate than estimated based on fossil record and *Picea–Pinus* divergence; (c) nonequilibrium population history reducing the long-term N_e ; and (d) effect of linked selection permeating most of the genic areas, thus reducing genetic diversity (Charlesworth, Morgan, & Charlesworth, 1993; Maynard Smith & Haigh, 1974).

These potential explanations do not exclude each other. It is likely that offspring number of *P. sylvestris* does not follow the assumed Poisson distribution with mean number of offspring = 1. Mature *P. sylvestris* trees have high fecundity, and thus, single trees have a potential to make a large contribution to next generation. This can lead to Moran model like multiple-merger population coalescent processes, reduced diversity in relation to population census size, and also to an excess of rare alleles due to star-shaped gene genealogies (Eldon & Wakeley, 2006). In seed orchard conditions, significant variation in offspring number has indeed been identified among genotypes (Gömöry, Bruchánik, & Longauer, 2003; Gömöry, Bruchanik, & Paule, 2000; Kang, Bila, Harju, & Lindgren, 2003; Savolainen, Kärkkäinen, Harju, Nikkanen, & Rusanen, 1993). In fully natural conditions, this variance is expected to be even larger. The AFS indicates nonequilibrium population history (Kujala & Savolainen, 2012;

Pyhäjärvi et al., 2007), which also reduces the amount of genetic diversity. Linked selection has been shown to affect a wide spectrum of species (Charlesworth & Charlesworth, 2018), and evaluating its importance in *P. sylvestris* is one of the major evolutionary questions that more genomic data will help us to tackle. Grivet et al. (2017) observed both high efficiency of purifying selection and high rate of positive selection in *P. sylvestris* and *Pinus pinaster*, in accordance with gymnosperms in general (De La Torre et al., 2017), supporting linked selection as a likely explanation for low genetic diversity in *P. sylvestris* genic areas. Evaluation of the overall effect of linked selection on genetic diversity requires a dense genetic map combined with the physical map because a correlation between nucleotide diversity and recombination rate is expected. Current estimates suggest that LD in *P. sylvestris* in general decays fast, often within few hundred bp (however, see detailed discussion on LD below), which would predict local effects of linked selection nearby genic regions, as observed, for example, in maize that also has low LD (Beissinger et al., 2016). However, the partial selfing may generate some opportunity for linked selection, but this has not yet been studied.

Note that different genetic estimates and measures of nucleotide diversity may have apparent discrepancy due to scale. Some estimates such as nucleotide diversity are estimated at base-pair resolution. However, other observations, such as lethal equivalents, reflect whole genome-level phenomena. The genomewide mutation rate to deleterious alleles can be high even with low per base-pair mutation rate level, if the mutational target size is large.

The recombination rate per bp (c) and extent of LD are critical factors in breeding as they determine how selection on a subset of loci will affect other nearby loci and essentially determines the genomic resolution of many breeding efforts. In an equilibrium situation, LD depends on both c and N_e and the population-level recombination parameter $\rho = 4N_e c$ be estimated from LD patterns. Further, c can be independently estimated with genetic mapping. In a practical context, it is good to remember that c affects the precision of, for example, QTL and other genetic mapping efforts, whereas ρ is more significant for association studies.

As mentioned above, *P. sylvestris* appears to have relatively low LD, when measured by r^2 , squared correlation coefficient, extending the level above 0.2 often only just few hundreds of base pairs (Kujala & Savolainen, 2012; Pyhäjärvi et al., 2007; Tyrmi et al., 2019). The extent of LD measured with r^2 is useful for many applications, such as genomic prediction, because it directly informs about the power of a locus to predict the allelic state of another locus (Hahn, 2018). However, as all measures of LD, it is dependent on the marginal allele frequencies. When there are many low-frequency SNPs, as in *P. sylvestris*, r^2 tends to be low not only due to high recombination but also due to low allele frequencies. *P. sylvestris* LD measured as $|D'|$ (LD measure scaled by its maximum value given the allele frequencies) that is more informative about the recombination rate suggests more extensive LD than the value from the overall r^2 decay (Kujala & Savolainen, 2012).

As in most species, LD patterns in conifers are probably heterogeneous among different genomic regions (Pavy, Namroud, Gagnon,

Isabel, & Bousquet, 2012). Despite the general trend of rapid LD decay (Pyhäjärvi et al., 2007; Wachowiak, Balk, & Savolainen, 2009), there is considerable variation along the genome in *P. sylvestris* at the gene level LD (Kujala & Savolainen, 2012; Pyhäjärvi, Kujala, & Savolainen, 2011). For example, Pyhäjärvi et al. (2011) found that several allozyme coding loci have strong LD, not showing signs of decay even over a 12-kbp region. Also in *P. taeda*, earlier data suggested that LD decays rapidly (Brown, Gill, Kuntz, Langley, & Neale, 2004), but some recent work suggests large variation in the extent of LD in the genome (Lu et al., 2016, see, however, Acosta et al., 2019).

The overall recombination rate in conifers, estimated based on genetic maps, is in general low (Jaramillo-Correa, Verdú, & González-Martínez, 2010) and the same applies to *P. sylvestris*. The estimated map length is 1,500 cM (Komulainen et al., 2003) and with the genome size 22×10^9 bp results in recombination rate of 0.07 cM/Mb or $c = 0.7 \times 10^{-9}$ per bp per generation. Thus, assuming N_e 38,000–230,000 obtained from nucleotide diversity data yields ρ estimates 0.0001–0.0006, whereas some ρ -estimates obtained from the nucleotide diversity data are much higher ranging from 0 to 0.04 (Kujala & Savolainen, 2012; Pyhäjärvi et al., 2007, 2011; Wachowiak et al., 2009). The apparent discrepancy of low LD and low recombination rate can be explained by low minor allele frequencies, relatively high N_e and potential for low recombination and extensive LD in intergenic regions. Currently, LD estimates of *P. sylvestris* are available within genes and between pairs of coding areas. Evidence for LD in intergenic areas is rare, and studies at an intermediate range are missing in *P. sylvestris*. A study on *Cryptomeria japonica* has indeed shown that noncoding regions of conifer genomes can harbor extensive LD (Moritsuka et al., 2012). Better genome assemblies combined with extensive resequencing efforts are required to get a fuller picture of LD across *P. sylvestris* genome. Long read sequencing combined with, for example, optical and genetic mapping will allow identifying regions where physical and genetic distances have most discrepancies. Identifying these regions is important for breeding and understanding adaptive variation as large fragments of genome are dragged along when responding to selection.

4 | POPULATION STRUCTURE, HISTORY, AND ALLELE FREQUENCY DISTRIBUTION

Pinus sylvestris has efficient wind pollination with potential for very long-distance pollen dispersal. In addition, female flowering precedes the male flowering by two to five days (Koski, 1970; Varis, Pakkanen, Galofré, & Pulkkinen, 2009). *Pinus sylvestris* pollen dispersal distribution has a “fat-tailed” leptokurtic shape allowing sporadic long-distance dispersal events. Most pollen comes from nearby sources (Koski, 1970; Muona & Harju, 1989; Torimaru, Wang, Fries, Andersson, & Lindgren, 2009), but nonlocal airborne germinable pollen is often observed during the female flowering, and northern populations receive some southern pollen potentially from hundreds of kilometers away before the local pollen is available (Varis et al., 2009). In Robledo-Arnuncio and Gil (2005), 4.3% of the pollen came

outside the isolated *P. sylvestris* stand despite an estimated average pollen dispersal distance of only 135 m (Robledo-Arnuncio & Gil, 2005). Even rare long-distance dispersal can have an important role in homogenizing the distribution of genetic variation and result in, for example, suboptimal phenotypic variation. Seeds are also dispersed by wind, but not as extensively as pollen, distances peaking in <10 m (Kellomäki, Hänninen, Kolström, Kotisaari, & Pukkala, 1987).

The dispersal biology is reflected in the geographic distribution of genetic diversity. In nuclear genes that are dispersed by both pollen and seeds, the genetic differentiation is consistently low ($F_{ST} = 0.02$) in the more continuous part of the distribution (Karhu et al., 1996; Kujala & Savolainen, 2012; Prus-Głowacki et al., 2012; Pyhäjärvi et al., 2007). Maternally inherited mitochondrial markers have a more restricted geographic distribution of alleles ($G_{ST} = 0.66$; Naydenov, Senneville, Beaulieu, Tremblay, & Bousquet, 2007; Pyhäjärvi, Salmela, & Savolainen, 2008). The eastern part of the distribution is less studied in terms of nuclear markers, but isozyme studies imply that genetic differentiation is also low (Dvornyk, 2001; Goncharenko, Silin, & Padutov, 1994). As in many other species, subtle geographic structure can be observed at the range margins and even in the main distribution when a large number of nuclear loci are observed (Kujala & Savolainen, 2012; Tyrmi et al., 2019; Wachowiak et al., 2011). In summary, the nuclear polymorphisms and continuous distribution indicate a lack of actual discrete populations, and in many analyses, *P. sylvestris* within most of Europe can be treated as a single panmictic population.

Pinus sylvestris distribution is continuous in the north, whereas in the southern margin, the distribution consists more of fragmented and isolated populations. Mediterranean peninsulas and Turkish populations harbor mitochondrial haplotypes that have rarely or never been observed outside these isolated populations (Cheddadi et al., 2006; Naydenov et al., 2007; Pyhäjärvi et al., 2008; Sinclair, Morman, & Ennos, 1999; Soranzo, Alia, Provan, & Powell, 2000; Wójcikiewicz, Cavers, & Wachowiak, 2016). This reflects the existence of limited seed dispersal and Mediterranean refugia during the Last Glacial Maximum (LGM) and still ongoing postglacial changes in suitable habitats, but also contemporary land use and level of forested areas in general. In addition to contemporary gene flow, part of the low nuclear differentiation in higher latitudes could be explained by colonization process combined with a long juvenile stage (Austerlitz, Mariette, Machon, Gouyon, & Godelle, 2000).

Analysis of population genetic structure, measured as F_{ST} or inferred using STRUCTURE type approaches (Falush, Stephens, & Pritchard, 2003), is based on model of discrete populations. In reality, given the dispersal biology of *P. sylvestris*, the among-population covariance structure is probably more of isolation-by-distance (IBD) type. Bradburd, Coop, and Ralph (2018) presented a promising new method to account for IBD patterns in genetic population structure analysis. When applied to nuclear *P. sylvestris* data, the spatial model with IBD has a better predictive accuracy than nonspatial (cluster) model (Tyrmi et al., 2019). The spatial model identifies a genetic component that is only present in an isolated Spanish population, consistent with previous work (Figure 3). However, in the spatial model the genetic makeup of the rest of the sampling sites remains continuous

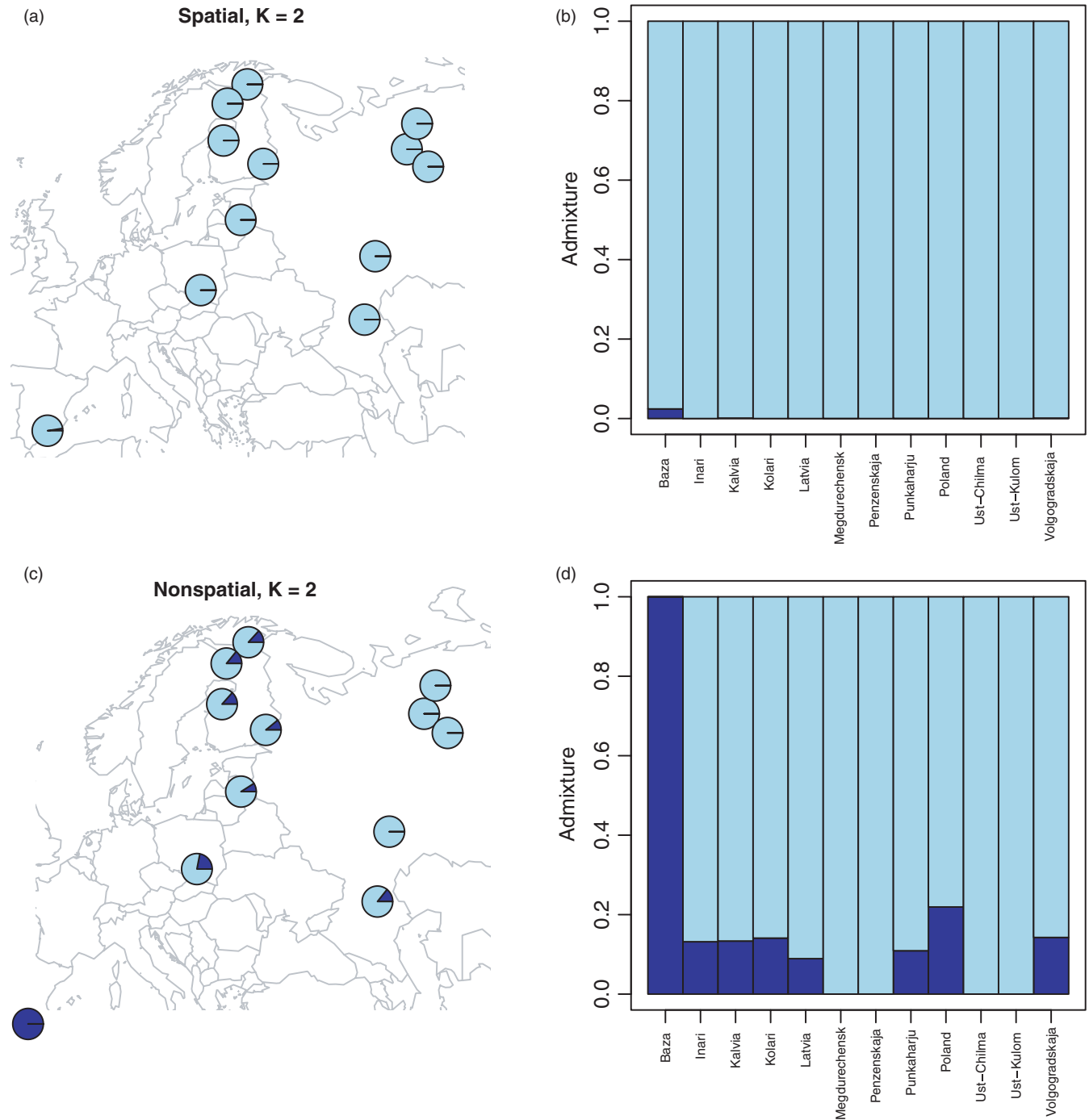


FIGURE 3 Application of spatial and nonspatial population structure model (Bradburd et al., 2018) with $K = 2$ to *Pinus sylvestris* exome capture data from (Tyrmi et al., 2019). Geographic distribution of layer contributions for each population under spatial (a) and nonspatial models (c). The proportion of layer contributions in each population for spatial (b) and nonspatial models (d). Sampling locations are ordered according to longitudes

in contrast to the nonspatial model that assigns a proportion of western populations to the Spanish-type cluster (Figure 3). These results support including IBD-type spatial patterns in further analyses affected by population structure of *P. sylvestris* such as GWAS.

Nucleotide diversity studies have revealed a nonequilibrium pattern in the distribution of allele frequencies of *P. sylvestris* throughout its distribution, which is a common observation in forest trees

(Figure 4; Heuertz et al., 2006; Ingvarsson, 2008; Mosca et al., 2012; Zhou, Bawa, & Holliday, 2014). The observed excess of rare variants is consistent with a bottleneck whose timing is uncertain but much earlier than the LGM (Kujala & Savolainen, 2012; Pyhäjärvi et al., 2007).

As for many other conifers, the evolutionary timescales that influence both phenotypic and molecular genetic diversity in *P.*

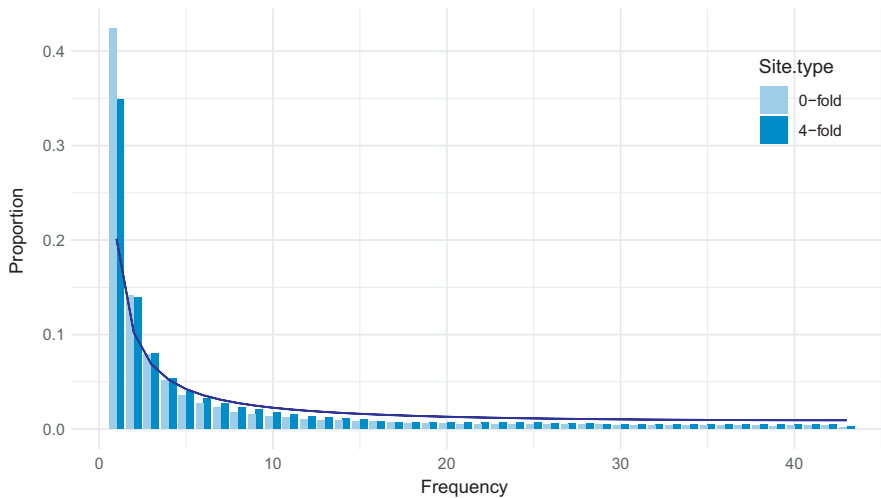


FIGURE 4 Example of an allele frequency spectrum in *Pinus sylvestris* (Tyrimi et al., 2019). Observed 4-fold and 0-fold degenerate sites are plotted as bars, and the expected values (Nordborg et al., 2005) under neutral equilibrium are indicated by line

sylvestris can be very extensive, potentially reaching millions of years (Pyhäjärvi et al., 2007). This is a combined property of its long generation time and large N_e . Therefore, observing the signatures of post-LGM population expansion would require large amount of data, exceeding the sample sizes used in most published studies. Simple coalescent simulations show that, for example, doubling the population size from 25×10^6 to 50×10^6 individuals during the past 20,000 years (1,000 generations) hardly shifts the distribution of Tajima's D from the equilibrium population expectations (Figure 5). The large N_e results in long expected coalescence times, and thus, most of the observed diversity reflects the time before LGM. However, recent expansion could partly explain the observed low nucleotide diversity (Figure 5). It is noteworthy that during adaptation in rapidly growing populations, the current population size governs the adaptive process, whereas overall molecular diversity is defined by long-term N_e , which can be considerably lower (Messer & Petrov, 2013). Larger sample sizes, allowed by more affordable sequencing, will provide a better resolution on addressing more recent demographic events as, for example, in Keinan and Clark (2012).

The skewed AFS also has practical consequences. The general utility and information content of a given polymorphism depends on its allele frequency. Rare alleles are informative only in a small number of populations and families. Common alleles are generally considered more useful in, for example, paternity analysis, breeding, genetic mapping, genomic prediction, and genomewide association analysis. However, in some cases, ignoring the rare alleles may lead to biased conclusions. It will result in biased estimates of diversity and may lead to ignoring important functional variation (De La Torre et al., 2017, 2019; Fahrenkrog et al., 2017; Manolio et al., 2009).

5 | ADAPTIVE VARIATION AT TRAIT AND MOLECULAR LEVEL

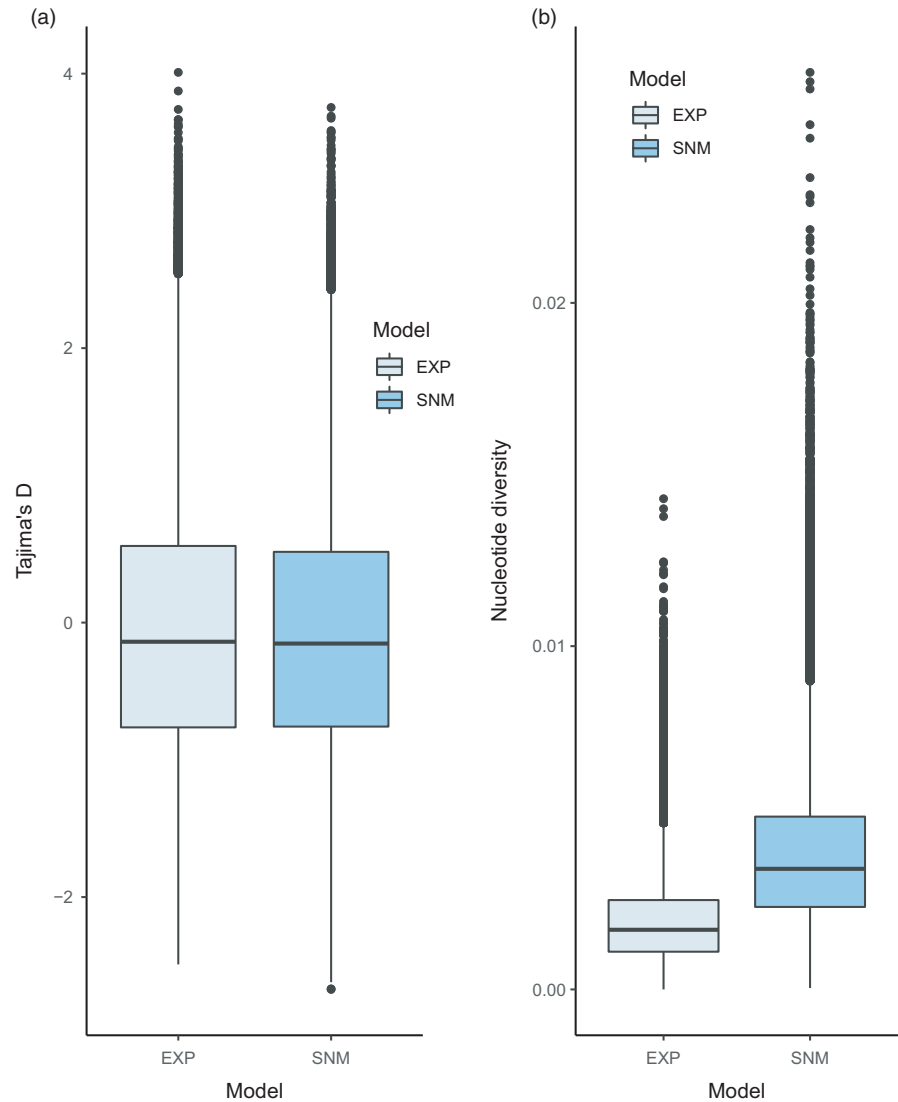
The extent of local adaptation and distribution of adaptive variation among geographic areas, genomes, and individuals is a core question in evolutionary genetics and also has impacts on conservation and

breeding efforts. The ultimate proof of local adaptation is the superior fitness of a local population in comparison with nonlocal populations (Kawecki & Ebert, 2004). So far, the strongest evidence for local adaptation in *P. sylvestris* has been obtained at the phenotypic level, but modern tools contribute to connecting phenotypic with molecular variation. Many of the potentially adaptive traits display a continuous, nearly normal within-population distribution indicative of polygenic basis of inheritance (Mather, 1943). Phenotypic variation is often also continuous at a geographic scale so that traits are correlated with, for example, altitude or latitude in a linear, clinal manner, caused by an interplay of selective gradient and gene flow in continuous space (Huxley, 1938). While many other conifer species have also been extensively characterized in terms of adaptive phenotypic variation, *P. sylvestris* is among the few for which extremely little population structure at genomewide level within a very large continuous distribution is coupled with high Q_{ST} values, measure of phenotypic differentiation among populations. These properties provide a relatively straightforward theoretical setup to study the dynamics of adaptive variation (Adrión, Hahn, & Cooper, 2015).

Barton (1999) suggests that a part of the loci affecting the clinal phenotypic variation should form strong allele frequency clines by the time population is approaching an equilibrium. Another suggestion emphasizes selection on favorable allele combinations across adaptive loci (allelic covariation) with only weak selection pressure on individual loci, instead of notable allele frequency clines. This could be especially important in the early phases of selection (Latta, 1998; Le Corre & Kremer, 2003, 2012). Further, phenotypes could well be genetically redundant, and allelic effects, transient (Barghi et al., 2019; Yeaman, 2015).

Over decades, *P. sylvestris* has been extensively characterized in terms of possible adaptive trait variation. In a comparison of 27 European conifer species, *P. sylvestris* had one of the highest Q_{ST} values on height increment, bud flush, and bud set (Alberto et al., 2013; see also Savolainen et al., 2007). Extensive phenotypic variation has been observed, for example, in phenology (Beuker, 1994; Clapham, Ekberg, Eriksson, Norell, & Vince-Prue, 2002; Kujala et al., 2017; Mikola, 1982; Notivol, García-Gil, Alia,

FIGURE 5 Tajima's D (a) and nucleotide diversity (b) distributions for constant size (SNM) and exponentially (EXP) growing population. Coalescent simulations were conducted with Coala package in R (Staab & Metzler, 2016) with the following parameters. SNM: $N_0 = 50 \times 10^6$, $\mu = 0.004/\text{bp}$ and locus length 1,000 bp and sample size 50. EXP: as in SNM but exponential growth started at time 0.00002



& Savolainen, 2007; Salmela, Cavers, Cottrell, Iason, & Ennos, 2013), cold tolerance (Hurme, Repo, Savolainen, & Pääkkönen, 1997), drought (Palmroth et al., 1999; Semerci et al., 2017), waterlogging stress (Donnelly, Cavers, Cottrell, & Ennos, 2018), root (Zadworny, McCormack, Mucha, Reich, & Oleksyn, 2016), seed (Reich, Oleksyn, & Tjoelker, 1994) and needle (Donnelly, Cavers, Cottrell, & Ennos, 2016) characteristics, and carbohydrate and nutrient dynamics (Oleksyn, Reich, Zytowskiak, Karolewski, & Tjoelker, 2003; Oleksyn, Zytowskiak, Karolewski, Reich, & Tjoelker, 2000). For example, northern populations are more cold tolerant, have earlier growth cessation, and grow less during the growing season, consistent with the gradient in environmental conditions and local adaptation.

Wealth of data on survival and growth serving as proxies for fitness exists in provenance trials (reviewed by Langlet (1971)). Savolainen et al. (2007) used such data to identify local adaptation. Comparison of fitness of transferred populations to the fitness of local populations based on relative survival and height showed, for example, that *P. sylvestris* populations in central Sweden are locally

adapted. This implies that tree populations have obviously had time to evolve to match the new habitats exposed by the retreating continental ice retreat (Davis & Shaw, 2001). Rapid adaptation is concordant with the theoretical expectation that selection is efficient in large populations.

The molecular genetic basis of the adaptive clinal variation can be identified with two basic approaches: association mapping that identifies genetic polymorphisms correlated with a given phenotypic variation, or by methods that rely solely on genotypic data. Association studies, especially when both the between-population and within-population variation can be addressed, can reveal important polymorphisms underlying adaptation. At the same time, the clinal setup poses a challenge for controlling population structure when it occurs along the same environmental gradient. Kujala et al. (2017) address this problem with latitudinal variation in timing of bud set in the first-year seedlings of *P. sylvestris*. In this Bayesian approach, the presence of a within-population association signal is required across different populations to exclude spurious associations. Other new promising methods, akin to Q_{ST} - F_{ST} comparisons,

for identification of polygenic adaptation rely on principal components of relatedness derived from genomic data and phenotype data obtained from common garden studies (Berg & Coop, 2014; Josephs, Berg, Ross-Ibarra, & Coop, 2019, see however Berg et al., 2019). This type of analysis should also be now feasible in *P. sylvestris*, utilizing one of the available high-throughput genotyping methods. The benefit of the Josephs et al. (2019) method is that it does not require a priori grouping of individuals into discrete populations and may be another solution to some of the analytical problems deriving from overcorrecting the population structure.

In *P. sylvestris*, association methods have been used to search for variants related to clinal variation in timing of bud set (Kujala et al., 2017). Most marker effects were small (<2 days), in line with most other genotype–phenotype association studies on growth cessation-related traits in trees (Holliday, Ritland, & Aitken, 2010; Ma, Hall, St. Onge, Jansson, & Ingvarsson, 2010; Mahony et al., 2019; Olson et al., 2013; Prunier et al., 2013; Yeaman et al., 2016). Interestingly, in the *P. sylvestris* study, different markers were associated in northern and central European populations, suggesting genetic heterogeneity within this trait. While some sharing of the associated loci with other species has been found, the between-species molecular convergence will be addressed in more detail in the upcoming association studies with higher genome coverage.

Divergence outlier methods (Beaumont & Nicholson, 1996; Excoffier, Hofer, & Foll, 2009; Foll & Gaggiotti, 2008) and deviations from the expected AFS have been used to identify loci indicating local adaptation or directional selection. However, inference of the effect of natural selection in *P. sylvestris* has been complicated by potential nonequilibrium demographic history. This has been accommodated by contrasting observed patterns of presumably neutral variation to a set of loci potentially under selection (Wachowiak et al., 2009). Another approach has been to build a demographic background model and derive expected patterns of neutral variation from that with coalescent simulations (Kujala & Savolainen, 2012; Pyhäjärvi et al., 2007). Even though most of these studies have been conducted on candidate genes, the number of outliers has usually been small (Kujala & Savolainen, 2012; Pyhäjärvi et al., 2007; Tyrmi et al., 2019; Wachowiak et al., 2009). However, despite moderate signals of selection at individual loci, Grivet et al. (2017) concluded that *P. sylvestris*, along with *P. pinaster*, have in general slightly higher substitution rates of adaptive mutations in comparison with many other plant species.

Genomes also carry evidence of very long-term selection, for example, in d_N/d_S ratios, as also identified in *P. sylvestris* (Palme, Pyhäjärvi, Wachowiak, & Savolainen, 2009; Palmé, Wright, & Savolainen, 2008). Identifying these kinds of “long-term” selected loci is not necessarily helpful in current breeding efforts, but they will inform about the long-term evolutionary forces shaping molecular patterns of, for example, divergence and transferability of the findings across species (Yeaman et al., 2016).

Very few nucleotide polymorphisms so far have indicated a clinal allele frequency pattern (Kujala & Savolainen, 2012; Tyrmi et al., 2019), and the overall lack of population-level adaptive signal at molecular level is in strong contrast with the observed patterns of

clinal phenotypic variation. This observation is consistent with the allelic covariation hypothesis, which would also explain the scarcity of other molecular level signal of adaptation. Alternatively, the clinal patterns could reside in yet-unexplored genes or their regulatory regions. In the similarly adapted *P. abies*, some examples of latitudinal allele frequency variation have been found (Chen et al., 2012). A recent study on *Pinus contorta* shows that joint multivariate analysis of environment with gene co-expression networks is a promising avenue to better grasp the architecture of polygenic clinal adaptation because in reality populations are adapting to different environmental dimensions at the same time, especially in complex landscapes (Lotterhos, Yeaman, Degner, Aitken, & Hodgins, 2018).

6 | GENOME SIZE AND ARCHITECTURE

A major aspect of *P. sylvestris* genetics is its large genome size, 22×10^9 bp (Bennett & Leitch, 2012). Assuming the composition is similar to the closely related *Pinus taeda* reference genome (Wegrzyn et al., 2014), only about 0.2% of its genome consists of coding regions. Pinus genomes have high repetitive content, extremely long introns (Stival Sena et al., 2014), large gene families, and potentially more than 50,000 genes (Stevens et al., 2016; Wegrzyn et al., 2014). A recent de novo transcriptome also indicates high complexity with more than 1.2×10^6 distinct transcripts observed when multiple tissues and genotypes are combined (Ojeda Alayon et al., 2018).

The annotation of pine reference genomes is still far from complete (Wegrzyn et al., in the same issue). Current *P. taeda* (v. 2.01) and *P. lambertiana* annotations (Box) reach at most 58% gene space completeness (Stevens et al., 2016) in terms of essential single-copy genes in Embryophyta (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015). In contrast, more than 80% of the same essential genes are transcribed, for example, in the *P. sylvestris* tissues (Ojeda Alayon et al., 2018). The discrepancy is caused by still fragmented assemblies, extremely long introns, intronless genes, pseudogenes, and the inability of annotation algorithms to identify conifer genes (Neale et al., 2014; Nystedt et al., 2013; Stevens et al., 2016; Stival Sena et al., 2014; Wegrzyn et al., 2014). The incompleteness can cause biased results in downstream genetic analysis. For example, when read mapping of *P. sylvestris* is conducted using *P. taeda* as a reference, assuming polymorphisms in nonannotated sites to be intergenic or neutral may lead to biased estimates of genetic diversity or prevalence of natural selection.

In addition to protein coding regions, conifers with large genomes may have relatively large functional space compared to, for example, angiosperm trees. There is likely a wealth of regulatory variation outside the coding region. Recent theory on the role of genome size suggests that large genomes have more potential for adaptive variation in noncoding region than small genomes (Mei, Stetter, Gates, Stitzer, & Ross-Ibarra, 2018) and that the difference in mutational target can affect the expected dynamics of adaptation (Höllinger, Pennings, & Hermisson, 2019), which may partly explain why strong

Box 1 Current genomic resources relevant for *Pinus sylvestris*

Draft genome of *P. sylvestris* (ftp://plantgenie.org/Data/ConGenIE/Picea_abies/v1.0/FASTA/GenomeAssemblies/Psylvestris_1Kbp_fa.gz, Nystedt et al. (2013)) contains 881,136 > 1 kbp unannotated contigs. 454 and Illumina raw data are downloadable in ENA database (project ERP002572).

De novo *P. sylvestris* **transcriptomes** are available for needles, pollen, vegetative buds, phloem, and various developmental stages of embryos and megagametophytes (Höllbacher, Schmitt, Hofer, Ferreira, & Lackner, 2017; Merino et al., 2016; Ojeda Alayon et al., 2018; Wachowiak, Trivedi, Perry, & Cavers, 2015). The haploid-guided Trinity assembly combining data from multiple tissues and genotypes has the highest completeness (https://pinus_sylvestris_transcriptome_public_data.object.pouta.csc.fi/Pinus_sylvestris_transcriptomes_Ojeda_2018.tar).

TreeGenes (<https://treegenesdb.org>, Wegrzyn et al., 2008) provides access to several (including *P. taeda* and *P. lambertiana*) tree genomes, annotations, transcriptomes, and other resources and tools. It also provides *P. sylvestris* data obtained from NCBI for ESTs, cDNA, and TSA (Transcriptome Shotgun Assembly) databases. Somewhat overlapping with ConGenie.

ConGenie (<http://congenie.org>, Nystedt et al.) allows access to multiple conifer reference sequences, annotated genes, transcriptomes, and ESTs.

PIER database (Pine Interspersed Element Resource, https://www.treegenesdb.org/FTP/Genomes/Pita/Repeats/Pine_Interspersed_Repeats_%28PIER%29_v1.0.fa, Wegrzyn et al.) contains sequence information on 19,194 repeats first identified in *P. taeda*.

Gymno PLAZA 1.0 (<https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/>, Proost et al., 2015) is a comparative genomics platform for 11 conifer species, including *P. sylvestris*. Information on gene families and orthological relationships among genes and species.

A set of **single-copy genes** in gymnosperms, including *P. sylvestris* (http://bioinformatics.psb.ugent.be/supplementary_data/zheli/gbe/, Li et al., 2017).

cpDNA genome and fragmented mtDNA genome sequences (<https://www.ebi.ac.uk/ena/data/view/PRJEB18435>, Donnelly et al., 2016, <https://www.ncbi.nlm.nih.gov/nuccore/JN854158.1>, Parks et al., 2012) can be used, for example, to identify sequencing reads originating from organellar genomes.

STAPLER Pipeline maker (Tyrmí, 2018) has been developed in conjunction with *P. sylvestris* exome capture workflow and functions with, for example, large *P. taeda* reference genome.

signals of adaptive variation in *P. sylvestris* genetic polymorphism (mostly targeting coding regions) have not been identified. Even though the genomic resources at the moment for *P. sylvestris* are limited, new techniques, such as ATAC-seq, allow targeting the open chromatin and therefore access the regulatory regions of even large plant genomes (Buenrostro, Wu, Chang, & Greenleaf, 2015; Maher et al., 2018). The mutation rate per bp per year in *Pinus* is not extremely high (Buschiazzi et al., 2012), but the large functional space may also explain the large number of lethal equivalents (Koelewijn et al., 1999; Lande et al., 1994; Williams, 2009). The findings of both Avia, Kärkkäinen, Lagercrantz, and Savolainen (2014) and Alakärppä et al. (2018) on population-specific expression patterns in *P. sylvestris* genes related to photoperiod circadian clock functions provide evidence of the importance of differential regulation. Interestingly, Alakärppä et al. (2018) also identified population-specific expression patterns of DNA methyltransferase (DNMT) genes, encouraging further studies on expression regulations role in local adaptation.

Recently, the effects of transposable elements (TEs) on trait and adaptive variation via regulating gene expression have become evident across a wide variety of species (Chuong, Elde, & Feschotte, 2017). In large genomes filled with TEs, it is likely that adaptive mutations happen by TE movement and consequent epigenetic changes in regulatory regions (Mei et al., 2018).

Transposable element sequences make up 80% of *Pinus* genomes in general (Stevens et al., 2016; Wegrzyn et al., 2014). Even if TEs are generally thought to be old and inactive, some are actively transcribed in certain tissues and conditions in *Pinus* (Cañas et al., 2017; Voronova, Belevich, Jansons, & Rungis, 2014). Similarly to other *Pinus*, *IFG-7a_PTa* is the most abundant LTR family in *P. sylvestris* (Voronova, Belevich, Korica, & Rungis, 2017). Some families, such as *PtAngelina*, are especially abundant in *P. sylvestris* (Voronova et al., 2017), and several low copy retrotransposon family abundancies vary among *P. sylvestris* individuals. The intraspecific variation in TE abundancies indicates that they also have potential to relate to other intraspecific variation in, for example, adaptive traits or levels of inbreeding depression. Even though whole genome resequencing is still relatively expensive, a combination of low depth DNA sequencing and RNAseq can be used in more detailed studies on relative abundances of TEs and their activity across tissues, genotypes, and geographic regions.

One promising method for accessing genomic diversity in *P. sylvestris* is targeted sequencing where custom-designed baits are used to target the regions of interest, such as exons in the genome. Baits can be designed by combining transcriptome data with information from reference genomes (Rellstab, Dauphin, Zoller, Brodbeck, & Gugerli, 2019; Suren et al., 2016). Tyrmí et al. 2019 have applied

the method successfully in *P. sylvestris* and obtained good-quality SNP data from 80,000 SNPs and 109 individuals with 30 Illumina HiSeq 2000 lanes. Even this approach has its challenges. With larger genomes, the off-target sequence and paralog regions are harder to exclude (Rellstab et al., 2019). However, applying filtering for excess heterozygosity and deviations from expected read ratios have been successfully used to identify paralog mapping (different gene copies mapping to a single reference locus) in, for example, *Pinus cembra* (McKinney, Waples, Seeb, & Seeb, 2017; Rellstab et al., 2019).

As already done in reference genome sequencing and transcriptome sequencing (Ojeda Alayon et al., 2018; Stevens et al., 2016), it is advisable to utilize sequence data from the haploid megagametophyte tissue in the quality control, as no heterozygotes are expected, and paralog problems can be avoided. Heterozygous calls are not expected in haploid material, and they can be used to remove regions with paralog mapping issues. Utilizing haploid tissue also facilitates genotype calling with comparatively smaller average depth (10×) as there is no need to identify heterozygotes reliably.

RNA sequencing can also be used to identify polymorphisms (Ojeda Alayon et al., 2018; Wachowiak et al., 2015). However, one should proceed with caution in making detailed estimates of genetic diversity and AFS based on RNA-seq due to, for example, allele-specific expression. Ojeda Alayon et al. (2018) provide multiple *P. sylvestris* reference de novo transcriptome assemblies of which some are more suitable for use in gene expression studies and others for population genetic analysis, SNP identification, and, for example, exome capture bait design.

Pinus taeda can be used as a mapping reference for *P. sylvestris*, as the species have a rather low estimated sequence divergence of 3% at silent sites (Grivet et al., 2017; Palmé et al., 2008). Compared to *P. sylvestris* transcriptome and the draft genome sequence (Box), it is probably the most comprehensive reference especially in the case when, for example, read data from targeted sequencing contain a lot of off-target sequence not present in the transcriptome. Mapping to a diverged reference is bound to lead to some extent of reference bias and, for example, exclusion of highly diverged, potentially evolutionarily important genomic regions (Kronenberg et al., 2018 for an example in primates). This effect is likely to be even more serious in noncoding areas and further complicates identifying adaptive regulatory alleles that may be especially important in species with large genomes (Mei et al., 2018). *P. sylvestris*-specific genome assembly would improve the quality of inference based on DNA sequence data and widen the available analytical repertoire (e.g., identification of copy number and structural variants, haplotype structure, and long-distance LD). Box is a collection of useful genomic resources and databases available for *P. sylvestris* at the moment.

7 | APPLICATION OF GENOMICS IN BREEDING AND CONSERVATION

Combining the phenotypic information from the phenotype and theory-based knowledge with modern genomic technologies provides

an opportunity for improved breeding methods and better predictions of *P. sylvestris* response to future environmental changes. Predicting responses to artificial or natural changes requires an understanding of both basic biological properties, evolutionary processes and genomic basis of adaptive trait variations, combined with practical aspects and reality of breeding and conservation activities. Two main methods, GWAS and genomic prediction, hold most potential to link genomic and phenotypic data together for application purposes.

GWAS uses large genomewide polymorphism and phenotypic data from typically large population samples to identify genomic variants associated with variation in a given trait and is based on tight LD between the marker and the causative locus (Risch & Merikangas, 1996; for trees, see Neale & Savolainen, 2004). So far, the association analyses in *P. sylvestris* have been based on a limited number of mostly candidate loci. Exome capture and transcriptome sequencing methods have now identified enough SNPs to enable design of efficient and affordable chip-based genotyping methods (Stephen Cavers, personal communication). Also, genotyping-by-sequencing (Elshire et al., 2011) and other restriction-based high-throughput methods facilitate genotyping of hundreds of individuals for large amounts of SNPs. However, these methods have most power to detect SNPs in the genic, nonrepetitive part of the genome. If genetic basis of trait variation is highly polygenic and outside coding region, whole genome resequencing and targeted sequencing of active chromatin are the methods that offer better resolution. Both of these methods are still economically and technically challenging for conifers.

In a typical association study, the effects of rare variants are considered extremely hard to capture due to power issues. The lower the frequency of the variant is, the more samples are needed to reach sufficient power given a certain effect size. As the number of tests grows, the significance threshold grows due to corrections for multiple testing (however, note multilocus approaches, e.g., Kujala et al. (2017)). However, when the AFS is very skewed, and a minor allele frequency cutoff used, much genomic information is lost with rare variants, and the effects (small or large) of low-frequency variants will not be evaluated at all (Auer & Lettre, 2015; Sazonovs & Barrett, 2018). The young variants are not efficiently tagged by older nearby variants due to the allele frequency difference (Auer & Lettre, 2015; Paulose, Hermisson, & Hallatschek, 2019; Sazonovs & Barrett, 2018; Wainschein et al., 2019). Family-based association methods, where the alleles that are rare in the general population are “amplified” within a given family (Laird & Lange, 2006), extreme phenotype sampling from the edges of the trait distribution (Guey et al., 2011; Li, Lewinger, Gauderman, Murcay, & Conti, 2011) or variance-component tests (Lee, Abecasis, Boehnke, & Lin, 2014; Wu et al., 2011) can be used to increase power and efficiency in case of rare variants.

Genomic prediction is based on the idea that a limited number of genetic markers can inform about areas of the genome responsible for trait variation. It relies on extensive LD between genetic markers and the causal polymorphism, typically not observed at the population-level samples. However, reduced N_e and family structure

in breeding populations can increase the extent of LD. The observed LD patterns and empirical results in other conifers suggest that genomic prediction is feasible in cases where selection takes place within families (Chen et al., 2018; Lenz et al., 2017). *P. sylvestris* breeding, for example, in Finland is conducted in such a setting and allows genomic prediction to be developed.

The low success of transferring genomic prediction models across different breeding populations is due to differences in the patterns of LD between populations. This limitation has been observed in trees (Isik, 2014), as well as in other species (Meuwissen, Hayes, & Goddard, 2016). A dual strategy could alleviate this problem; instead of using an array consisting only of anonymous genotypic markers, one could allow more emphasis on markers in functionally relevant genes (Meuwissen et al., 2016; Nicolae et al., 2010; Wang et al., 2018; Zhang et al., 2015, 2010), for example, from well-known pathways (such as pinosylvin biosynthesis; Lim et al., 2016). However, only those functional markers that have an impact on the variation across different populations would serve this purpose; markers with more locally distributed effects in contrary might enhance misinference. The above-cited methods can however help to improve the genomic predictions within populations, too. Polymorphisms indicated in association studies in conifers are expected to be very close to the causal loci due to low LD, and these markers could thus be also given higher weight in genomic prediction. The excess of rare alleles is a challenge but also an opportunity for *P. sylvestris* breeding. Recent studies have shown that rare alleles may account for a considerable proportion of phenotypic variation and that some of these rare alleles may be deleterious, maintained by mutation-selection balance (Wainschtein et al., 2019). There is definitely call for identifying and using rare variants also in plant breeding (Bernardo, 2016; Crossa et al., 2017). Further, they give highly accurate estimates of true genetic relationships and level of inbreeding (Eynard, Windig, Leroy, Van Binsbergen, & Calus, 2015). Jannink (2010) and Liu, Meuwissen, Sørensen, and Berg (2015) have shown that weighting low-frequency favorable alleles increases the long-term genetic gain in genomic prediction. The effect comes through reducing the odds of losing rare variation through genetic drift. This comes at the expense of short-term genetic gain, but maintaining variation in the breeding populations is of high importance, too.

In *P. sylvestris*, each family or other subsample will have their unique set of rare variants. They are likely to be missed by, for example, general purpose SNP chips and can only be identified by sequencing-based methods. Population stratification in rare alleles is likely to differ from the pattern observed with more common alleles as they often show more spatial clustering. Therefore, more rigorous methods to deal with the fine spatial structure are needed when working with rare variants (Mathieson & McVean, 2012). On the other hand, rare variants offer a way to learn more about the potential fine-scale spatial structure in *P. sylvestris*, a topic not well covered so far.

Warming temperatures are already causing range shifts both at the northern and southern distribution limits of species, including *P. sylvestris* (Dyderski, Paż, Frelich, & Jagodziński, 2018). This is

expected to increase growth rates of *P. sylvestris* in northern areas (Barring, Berlin, & Andersson Gull, 2017; Bergh et al., 2003; Kellomäki et al., 2018). However, for this growth increase to be realized, forest reproductive materials with appropriate adaptive characteristics must be used (Beuker, 1994; Persson, 1998; Persson & Beuker, 1997; Rehfeldt et al., 2002) taking into account the complicated joint effects of abiotic and biotic changes (Matias & Jump, 2012). In some other areas, growth and survival may be reduced (Reich & Oleksyn, 2008). New combinations of photoperiod and temperature create selection pressure to efficiently utilize the increased potential for growth by correct interpretation of climatic signals (Saikkonen et al., 2012). When the genetics and relationship of phenotypic variation to environmental variation are understood, it is possible to predict and evaluate responses (Kuparinen, Savolainen, & Schurr, 2010; Savolainen, Bokma, García-Gil, Komulainen, & Repo, 2004) and utilize this information, for example, in assisted migration design.

Along with the longer growth periods, also many forest disturbance effects are likely to take place. According to Seidl et al. (2017), temperature-related disturbances will be highest in the boreal biome, especially in coniferous forests. Drought effects will be significant in both southern and northern regions, flooding will likely be an increasing problem, and new pests and pathogens will be encountered. It should be of high priority to assess the levels of within-species and within-population variation for adaptive potential against the predicted disturbances, at both phenotypic and genomic levels. Identification of the adaptive genomic regions, and diversity within those regions, albeit not an easy task, should be seen as an important task. Genetic analyses can also inform us, for example, about the extent of correlation among adaptively important traits, deriving either from pleiotropic effects or tight LD (Aitken, Yeaman, Holliday, Wang, & Curtis-McLane, 2008; Savolainen et al., 2004). Further, genomic prediction may contribute to climate change responses helping meet goals of carbon neutrality (Cuny et al., 2015; Pan et al., 2011; Pool & Aquadro, 2007; Varho, Rautiainen, Peltonen, Niemi, & Ovaska, 2018) and alleviate associated disturbances (Alberto et al., 2013; Fady et al., 2016; Jansson et al., 2017; Savolainen et al., 2004; Seidl et al., 2017).

In summary, new genetic technologies and the molecular diversity they reveal—DNA polymorphisms, expression changes, structural, and repetitive element variation—are opening many venues to take a new look into questions identified by earlier *P. sylvestris* research and also applicable to many other forest trees. Bridging the artificial gap between quantitative, population, and evolutionary genetics is necessary for an in-depth understanding of causes and consequences of, for example, clinal adaptive variation, inbreeding depression, large genome size, and skewed AFS in *P. sylvestris*.

ACKNOWLEDGEMENTS

The authors would like to thank laboratory members of Brandvain and McGaugh at University of Minnesota and Katri Kärkkäinen at Natural Resources Institute Finland for helpful discussions throughout the manuscript preparation. Jaakko Tyrmä, Zhen Li and Jill Wegrzyn kindly shared data and discussed bioinformatic issues. The

work was funded by Academy of Finland (287431 and 293819 to TP, 309978 to STK, and 307582 GENOWOOD to OS) and European Union's Horizon 2020 research and innovation programme under grant agreement No 773383 (B4EST).

CONFLICT OF INTEREST

None declared.

DATA ACCESSIBILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Tanja Pyhäjärvi  <https://orcid.org/0000-0001-6958-5172>

Sonja T Kujala  <https://orcid.org/0000-0003-0949-6156>

Outi Savolainen  <https://orcid.org/0000-0001-9851-7945>

REFERENCES

- Acosta, J. J., Fahrenkrog, A. M., Neves, L. G., Resende, M. F. R., Dervinis, C., Davis, J. M., ... Kirst, M. (2019). Exome re-sequencing reveals evolutionary history, genomic diversity, and targets of selection in the conifers *Pinus taeda* and *Pinus elliottii*. *Genome Biology and Evolution*, 11(2), 508–520. <https://doi.org/10.1093/gbe/evz016>
- Adrion, J. R., Hahn, M. W., & Cooper, B. S. (2015). Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. *Trends in Genetics*, 31(8), 434–444. <https://doi.org/10.1016/j.tig.2015.05.006>
- Ahtikoski, A., Ojansuu, R., Haapanen, M., Hynynen, J., & Kärkkäinen, K. (2012). Financial performance of using genetically improved regeneration material of Scots pine (*Pinus sylvestris* L.) in Finland. *New Forests*, 43(3), 335–348. <https://doi.org/10.1007/s11056-011-9284-6>
- Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evolutionary Applications*, 1(1), 95–111. <https://doi.org/10.1179/1479096314Z.00000000018>
- Alakärppä, E., Salo, H. M., Valledor, L., Cañal, M. J., Häggman, H., & Vuosku, J. (2018). Natural variation of DNA methylation and gene expression may determine local adaptations of Scots pine populations. *Journal of Experimental Botany*, 69(21), 5293–5305. <https://doi.org/10.1093/jxb/ery292>
- Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., ... Savolainen, O. (2013). Potential for evolutionary responses to climate change – Evidence from tree populations. *Global Change Biology*, 19(6), 1645–1661. <https://doi.org/10.1111/gcb.12181>
- Auer, P. L., & Lettre, G. (2015). Rare variant association studies: Considerations, challenges and opportunities. *Genome Medicine*, 7(1), 1–11. <https://doi.org/10.1186/s13073-015-0138-2>
- Austerlitz, F., Mariette, S., Machon, N., Gouyon, P. H., & Godelle, B. (2000). Effects of colonization processes on genetic diversity: Differences between annual plants and tree species. *Genetics*, 154(3), 1309–1321.
- Avia, K., Kärkkäinen, K., Lagercrantz, U., & Savolainen, O. (2014). Association of FLOWERING LOCUS T/TERMINAL FLOWER 1-like gene FTL2 expression with growth rhythm in Scots pine (*Pinus sylvestris*). *The New Phytologist*, 204(1), 159–170. <https://doi.org/10.1111/nph.12901>
- Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F., Otte, K. A., ... Schlotterer, C. (2019). Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLOS Biology*, 17(2), 1–31. <https://doi.org/10.1371/journal.pbio.3000128>
- Bärring, L., Berlin, M., & Andersson Gull, B. (2017). Tailored climate indices for climate-proofing operational forestry applications in Sweden and Finland. *International Journal of Climatology*, 37(1), 123–142. <https://doi.org/10.1002/joc.4691>
- Barton, N. H. (1999). Clines in polygenic traits. *Genetical Research*, 74(3), 223–236. <https://doi.org/10.1017/S001667239900422X>
- Beaumont, M. A., & Nicholson, J. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1377), 1619–1626.
- Beissinger, T. M., Wang, L., Crosby, K., Durvasula, A., Hufford, M. B., & Ross-Ibarra, J. (2016). Recent demography drives changes in linked selection across the maize genome. *Nature Plants*, 2(7), 1–7. <https://doi.org/10.1038/nplants.2016.84>
- Bennett, M. D., & Leitch, I. J. (2012). Plant DNA C-values database (release 6.0, Dec. 2012).
- Berg, J. J., & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genetics*, 10(8), e1004412. <https://doi.org/10.1371/journal.pgen.1004412>
- Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., ... Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *eLife*, 8, 1–47. <https://doi.org/10.7554/eLife.39725>
- Bergh, J., Freeman, M., Sigurdsson, B., Kellomäki, S., Laitinen, K., Niinistö, S., ... Linder, S. (2003). Modelling the short-term effects of climate change on the productivity of selected tree species in Nordic countries. *Forest Ecology and Management*, 183(1–3), 327–340. [https://doi.org/10.1016/S0378-1127\(03\)00117-8](https://doi.org/10.1016/S0378-1127(03)00117-8)
- Bernardo, R. (2016). Bandwagons I, too, have known. *Theoretical and Applied Genetics*, 129(12), 2323–2332. <https://doi.org/10.1007/s00122-016-2772-5>
- Beuker, E. (1994). Adaptation to climatic changes of the timing of bud burst in populations of *Pinus sylvestris* L. and *Picea abies* (L.) Karst. *Tree Physiology*, 14(7–8–9), 961–970. <https://doi.org/10.1093/treephys/14.7-8-9.961>
- Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete Population genetic structure across space. *Genetics*, 210, 33–52. <https://doi.org/10.1534/genetics.118.301333>
- Brown, G. R., Gill, G. P., Kuntz, R. J., Langley, C. H., & Neale, D. B. (2004). Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences*, 101(42), 15255–15260. <https://doi.org/10.1073/pnas.0404231101>
- Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 109, 21.29.1–21.29.9. <https://doi.org/10.1002/0471142727.mb2129s109>
- Buschiazzo, E., Ritland, C., Bohlmann, J., & Ritland, K. (2012). Slow but not low: Genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evolutionary Biology*, 12(1), 8. <https://doi.org/10.1186/1471-2148-12-8>
- Cañas, R. A., Li, Z., Pascual, M. B., Castro-Rodríguez, V., Ávila, C., Sterck, L., ... Cánovas, F. M. (2017). The gene expression landscape of pine seedling tissues. *Plant Journal*, 91(6), 1064–1087. <https://doi.org/10.1111/tpj.13617>
- Charlesworth, B., & Charlesworth, D. (2018). Neutral variation in the context of selection. *Molecular Biology and Evolution*, 35(6), 1359–1361. <https://doi.org/10.1093/molbev/msy062>
- Charlesworth, B., Morgan, M., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134, 1289–1303. <https://doi.org/10.1021/bi00873a031>
- Cheddadi, R., Vendramin, G. G., Litt, T., Francois, L., Kageyama, M., Lorentz, S., ... Lunt, D. (2006). Imprints of glacial refugia in the modern

- genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography*, 15(3), 271–282. <https://doi.org/10.1111/j.1466-822X.2006.00226.x>
- Chen, J., Glémin, S., & Lascoux, M. (2017). Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution*, 34(6), 1417–1428. <https://doi.org/10.1093/molbev/msx088>
- Chen, J., Gyllenstrand, N., Zaina, G., Morgante, M., Bousquet, J., Eckert, A., ... Lascoux, M. (2012). Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics*, 191, 865–881. <https://doi.org/10.5061/dryad.82201.1>
- Chen, Z., Baisou, J., Pan, J., Karlsson, B., Gull, B., Westin, J., ... Wu, H. (2018). Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as a genotyping platform in Norway spruce. *BMC Genomics*, 19(1), 946.
- Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. *Nature Reviews Genetics*, 18(2), 71–86. <https://doi.org/10.1038/nrg.2016.139>
- Clapham, D. H., Ekberg, I., Eriksson, G., Norell, L., & Vince-Prue, D. (2002). Requirement for far-red light to maintain secondary needle extension growth in northern but not southern populations of *Pinus sylvestris* (Scots pine). *Physiologia Plantarum*, 114, 207–212. <https://doi.org/10.1034/j.1399-3054.2002.1140206.x>
- Conte, G. L., Hodgins, K. A., Yeaman, S., Degner, J. C., Aitken, S. N., Rieseberg, L. H., & Whitlock, M. C. (2017). Bioinformatically predicted deleterious mutations reveal complementation in the interior spruce hybrid complex. *BMC Genomics*, 18(1), 1–12. <https://doi.org/10.1186/s12864-017-4344-8>
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., ... Varshney, R. K. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science*, 22(11), 961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Cuny, H. E., Rathgeber, C. B. K., Frank, D., Fonti, P., Mäkinen, H., Prislan, P., ... Fournier, M. (2015). Woody biomass production lags stem-girth increase by over one month in coniferous forests. *Nature Plants*, 1(11), 15160. <https://doi.org/10.3389/fimmu.2013.00462>
- Davis, M. B., & Shaw, R. G. (2001). Range shifts and adaptive responses to Quaternary climate change. *Science*, 292(14), 1–8. <https://doi.org/10.1126/science.292.5517.673>
- De La Torre, A. R., Li, Z., Van De Peer, Y., & Ingvarsson, P. K. (2017). Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Molecular Biology and Evolution*, 34(6), 1363–1377. <https://doi.org/10.1093/molbev/msx069>
- De La Torre, A. R., Puiu, D., Crepeau, M. W., Stevens, K., Salzberg, S. L., Langley, C. H., & Neale, D. B. (2019). Genomic architecture of complex traits in loblolly pine. *New Phytologist*, 221(4), 1789–1801.
- Donnelly, K., Cavers, S., Cottrell, J. E., & Ennos, R. A. (2016). Genetic variation for needle traits in Scots pine (*Pinus sylvestris* L.). *Tree Genetics and Genomes*, 12(3), <https://doi.org/10.1007/s11295-016-1000-4>
- Donnelly, K., Cavers, S., Cottrell, J. E., & Ennos, R. A. (2018). Cryptic genetic variation and adaptation to waterlogging in Caledonian Scots pine, *Pinus sylvestris* L. *Ecology and Evolution*, 8(17), 8665–8675. <https://doi.org/10.1002/ece3.4389>
- Durrant, T. H., Rigo, D. D., & Caudullo, G. (2016). *Pinus sylvestris*. *European Atlas of Forest Tree Species*, 132–133. <https://doi.org/10.2788/038466>
- Dvornyk, V. (2001). Genetic variability and differentiation of geographically marginal Scots pine populations from Ukraine. *Silvae Genetica*, 50(2), 64–69.
- Dvornyk, V., Sirviö, A., Mikkonen, M., & Savolainen, O. (2002). Low nucleotide diversity at the pal1 locus in the widely distributed *Pinus sylvestris*. *Molecular Biology and Evolution*, 19(2), 179–188. <https://doi.org/10.1093/oxfordjournals.molbev.a004070>
- Dyderski, M. K., Paž, S., Frelich, L. E., & Jagodziński, A. M. (2018). How much does climate change threaten European forest tree species distributions? *Global Change Biology*, 24(3), 1150–1163. <https://doi.org/10.1111/gcb.13925>
- Eckert, A. J., Bower, A. D., Jermstad, K. D., Wegrzyn, J. L., Knaus, B. J., Syring, J. V., & Neale, D. B. (2013). Multilocus analyses reveal little evidence for lineage-wide adaptive evolution within major clades of soft pines (*Pinus* subgenus *Strobus*). *Molecular Ecology*, 22(22), 5635–5650. <https://doi.org/10.1111/mec.12514>
- Eiche, V. (1966). Cold damage and plant mortality in experimental provenance plantations with Scots pine in Northern Sweden. *Studia Forestalia Suecica*, 36, 1–219.
- Eldon, B., & Wakeley, J. (2006). Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*, 172(4), 2621–2633. <https://doi.org/10.1534/genetics.105.052175>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6(5), 1–10. <https://doi.org/10.1371/journal.pone.0019379>
- Engler, A. (1913). Einfluss der Provenienz des Samens auf die Eigenschaften der forstlichen Holzgewächse. Zweite Mitteilung. Mitt. Schweiz. Centralanst. f. das Forstl. Versuchsw., 9.
- Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, 103(4), 285–298. <https://doi.org/10.1038/hdy.2009.74>
- Eynard, S. E., Windig, J. J., Leroy, G., Van Binsbergen, R., & Calus, M. P. L. (2015). The effect of rare alleles on estimated genomic relationships from whole genome sequence data. *BMC Genetics*, 16(1), 24. <https://doi.org/10.1186/s12863-015-0185-0>
- Fady, B., Cottrell, J., Ackzell, L., Alía, R., Muys, B., Prada, A., & González-Martínez, S. C. (2016). Forests and global change: What can genetics contribute to the major forest management and policy challenges of the twenty-first century? *Regional Environmental Change*, 16(4), 927–939. <https://doi.org/10.1007/s10113-015-0843-9>
- Fahlvik, N., Ekö, P. M., & Pettersson, N. (2005). Influence of precommercial thinning grade on branch diameter and crown ratio in *Pinus sylvestris* in southern Sweden. *Scandinavian Journal of Forest Research*, 20(3), 243–251. <https://doi.org/10.1080/02827580510008266>
- Fahrenkrog, A. M., Neves, L. G., Resende, M. F. R., Vazquez, A. I., Campos, G., Dervinis, C., ... Kirst, M. (2017). Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. *New Phytologist*, 213(2), 799–811. others
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164(4), 1567–1587. <https://doi.org/10.1111/j.1471-8286.2007.01758.x>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180(2), 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Ford, G. A., McKeand, S. E., Jett, J. B., & Isik, F. (2014). Effects of inbreeding on growth and quality traits in loblolly pine. *Forest Science*, 61(3), 579–585. <https://doi.org/10.5849/forsci.13-185>
- García-Gil, M. R., Mikkonen, M., & Savolainen, O. (2003). Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Molecular Ecology*, 12(5), 1195–1206. <https://doi.org/10.1046/j.1365-294X.2003.01826.x>
- Gardner, M. (2013). *Pinus sylvestris*. The IUCN Red List of Threatened Species. Retrieved October 24, 2018, from <https://doi.org/10.2305/IUCN.UK.2013-1.RLTS.T42418A2978732.en>
- Giertych, M., & Matyas, C. (1991). *Genetics of Scots Pine*. Budapest, Hungary: Elsevier.
- Giertych, M., & Oleksyn, J. (1992). Studies on genetic variation in Scots pine (*Pinus sylvestris* L.) coordinated by IUFRO. *Silvae Genetica*, 41, 133–143.
- Gömöry, D., Bruchánik, R., & Longauer, R. (2003). Fertility variation and flowering asynchrony in *Pinus sylvestris*: Consequences for

- the genetic structure of progeny in seed orchards. *Forest Ecology and Management*, 174(1–3), 117–126. [https://doi.org/10.1016/S0378-1127\(02\)00031-2](https://doi.org/10.1016/S0378-1127(02)00031-2)
- Gömöry, D., Bruchanik, R., & Paule, L. (2000). Effective population number estimation of three Scots pine (*Pinus sylvestris* L.) seed orchards based on an integrated assessment of flowering, floral phenology, and seed orchard. *Forest Genetics*, 7(1), 65–75.
- Goncharenko, G. G., Silin, A. E., & Padutov, V. E. (1994). Allozyme variation in natural populations of Eurasian pines III Population structure, differentiation and gene flow in central and isolated populations of *Pinus sylvestris* L. in Eastern Europe and Siberia. *Silvae Genetica*, 43, 119–132.
- Grivet, D., Avia, K., Vaattovaara, A., Eckert, A. J., Neale, D. B., Savolainen, O., & González-Martínez, S. C. (2017). High rate of adaptive evolution in two widespread European pines. *Molecular Ecology*, 26(24), 6857–6870. <https://doi.org/10.1111/mec.14402>
- Guey, L. T., Kravic, J., Melander, O., Burt, N. P., Laramie, J. M., Lyssenko, V., ... Voight, B. F. (2011). Power in the phenotypic extremes: A simulation study of power in discovery and replication of rare variants. *Genetic Epidemiology*, 35(4), 236–246. <https://doi.org/10.1002/gepi.20572>
- Haapanen, M., Hynynen, J., Ruotsalainen, S., Siipilehto, J., & Kilpeläinen, M. L. (2016). Realised and projected gains in growth, quality and simulated yield of genetically improved Scots pine in southern Finland. *European Journal of Forest Research*, 135(6), 997–1009. <https://doi.org/10.1007/s10342-016-0989-0>
- Hahn, M. W. (2018). *Molecular population genetics*. Sunderland, MA: Oxford University Press.
- Hedrick, P. W., Hellsten, U., & Grattapaglia, D. (2016). Examining the cause of high inbreeding depression: Analysis of whole-genome sequence data in 28 selfed progeny of *Eucalyptus grandis*. *New Phytologist*, 209(2), 600–611. <https://doi.org/10.1111/nph.13639>
- Hedrick, P. W., & Muona, O. (1990). Linkage of viability genes to marker loci in selfing organisms. *Heredity*, 64(1), 67–72. <https://doi.org/10.1038/hdy.1990.8>
- Heikinheimo, O. (1950). Results of experiments on the geographical races of spruce and pine. *Communications Instituti Forestalis Fenniae*, 37(2), 1–36.
- Heuertz, M., De Paoli, E., Källman, T., Larsson, H., Jurman, I., Morgante, M., ... Gyllenstrand, N. (2006). Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics*, 174(4), 2095–2105. <https://doi.org/10.1534/genetics.106.065102>
- Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Rieseberg, L. H., & Aitken, S. N. (2016). Expression divergence is correlated with sequence evolution but not positive selection in conifers. *Molecular Biology and Evolution*, 33(6), msw032-. <https://doi.org/10.1093/molbev/msw032>
- Högberg, P., Nordgren, A., Buchmann, N., Taylor, A. F. S., Ekblad, A., Högberg, M. N., ... Read, D. J. (2001). Large-scale forest girdling shows that current photosynthesis drives soil respiration. *Nature*, 411, 789–793. <https://doi.org/10.1038/35081058>
- Höllbacher, B., Schmitt, A. O., Hofer, H., Ferreira, F., & Lackner, P. (2017). Identification of proteases and protease inhibitors in allergenic and non-allergenic pollen. *International Journal of Molecular Sciences*, 18(6), 1199. <https://doi.org/10.3390/ijms18061199>
- Holliday, J. A., Ritland, K., & Aitken, S. N. (2010). Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist*, 188(2), 501–514. <https://doi.org/10.1111/j.1469-8137.2010.03380.x>
- Höllinger, I., Pennings, P. S., & Hermisson, J. (2019). Polygenic adaptation: From sweeps to subtle frequency shifts. *PLOS Genetics*, 15(3), 1–26. <https://doi.org/10.1371/journal.pgen.1008035>
- Hurme, P., Repo, T., Savolainen, O., & Pääkkönen, T. (1997). Climatic adaptation of bud set and frost hardiness in Scots pine (*Pinus sylvestris*). *Canadian Journal of Forest Research*, 27, 716–723.
- Huxley, J. S. (1938). Clines: An auxiliary method in taxonomy. *Nature*, 142, 219–220.
- Ingvarsson, P. K. (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics*, 180(1), 329–340. <https://doi.org/10.1534/genetics.108.090431>
- Isik, F. (2014). Genomic selection in forest tree breeding: The concept and an outlook to the future. *New Forests*, 45(3), 379–401. <https://doi.org/10.1007/s11056-014-9422-z>
- Isik, F., Bartholomé, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., ... Bouffier, L. (2016). Genomic selection in maritime pine. *Plant Science*, 242, 108–119. <https://doi.org/10.1016/j.plantsci.2015.08.006>
- Jannink, J.-L. (2010). Dynamics of long-term genomic selection. *Genetics Selection Evolution*, 42(1), 35. <https://doi.org/10.1186/1297-9686-42-35>
- Jansson, G., Hansen, J. K., Haapanen, M., Kvaalen, H., & Steffenrem, A. (2017). The genetic and economic gains from forest tree breeding programmes in Scandinavia and Finland. *Scandinavian Journal of Forest Research*, 32(4), 273–286. <https://doi.org/10.1080/02827581.2016.1242770>
- Jaramillo-Correa, J. P., Verdú, M., & González-Martínez, S. C. (2010). The contribution of recombination to heterozygosity differs among plant evolutionary lineages and life-forms. *BMC Evolutionary Biology*, <https://doi.org/10.1186/1471-2148-10-22>
- Josephs, E. B., Berg, J. J., Ross-ibarra, J., & Coop, G. M. (2019). Detecting adaptive differentiation in structured populations with genomic data and common gardens. *Genetics Society of America*, 211(3), 989–1004.
- Kang, K.-S., Bila, A. D., Harju, A. M., & Lindgren, D. (2003). Estimation of fertility variation in forest tree populations. *Forestry*, 76(3), 329–344. <https://doi.org/10.1093/forestry/76.3.329>
- Karhu, A., Hurme, P., Karjalainen, M., Karvonen, P., Kärkkäinen, K., Neale, D., & Savolainen, O. (1996). Do molecular markers reflect patterns of differentiation in adaptive traits of conifers? *Theoretical and Applied Genetics*, 93(1), 215–221. <https://doi.org/10.1007/BF00225748>
- Kärkkäinen, K., Koski, V., & Savolainen, O. (1996). Geographical variation in the inbreeding depression of Scots pine. *Evolution*, 50(1), 111. <https://doi.org/10.2307/2410785>
- Kärkkäinen, K., & Savolainen, O. (1993). The degree of early inbreeding depression determines the selfing rate at the seed stage: Model and results from *Pinus sylvestris* (Scots pine). *Heredity*, 71(2), 160–166. <https://doi.org/10.1038/hdy.1993.120>
- Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, 7(12), 1225–1241. <https://doi.org/10.1111/j.1461-0248.2004.00684.x>
- Keinan, A., & Clark, A. G. (2012). Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*, 336(6042), 740–743. <https://doi.org/10.1126/science.333.6042.740>
- Kellomäki, S., Hänninen, H., Kolström, T., Kotisaari, A., & Pukkala, T. (1987). A tentative model for describing the effects of some regenerative processes on the properties of natural seedling stands. *Silva Fennica*, 21(1), 1–10.
- Kellomäki, S., Strandman, H., Heinonen, T., Asikainen, A., Venäläinen, A., & Peltola, H. (2018). Temporal and spatial change in diameter growth of boreal scots pine, Norway spruce, and birch under recent-generation (CMIP5) global climate model projections for the 21st century. *Forests*, 9(3), 5–7. <https://doi.org/10.3390/f9030118>
- Koelewijn, H. P., Koski, V., & Savolainen, O. (1999). Magnitude and timing of inbreeding depression in Scots pine (*Pinus sylvestris* L.). *Evolution*, 53(3), 758–768.
- Komulainen, P., Brown, G. R., Mikkonen, M., Karhu, A., García-Gil, M. R., O'Malley, D., ... Savolainen, O. (2003). Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. *Theoretical and Applied Genetics*, 107(4), 667–678. <https://doi.org/10.1007/s00122-003-1312-2>

- Koski, V. (1970). A study of pollen dispersal as a mechanism of gene flow in conifers. *Communications Instituti Forestalis Fenniae*, 70(4), 1–78.
- Koski, V. (1971). Embryonic lethals of *Picea abies* and *Pinus sylvestris*. *Communications Instituti Forestalis Fenniae*, 75(3), 1–30.
- Kronenberg, Z. N., Fiddes, I. T., Gordon, D., Murali, S., Cantsilieris, S., Meyerson, O. S., ... Eichler, E. E. (2018). High-resolution comparative analysis of great ape genomes. *Science*, 360(6393), eaar6343. <https://doi.org/10.1126/science.aar6343>
- Kujala, S. T., Knürr, T., Kärkkäinen, K., Neale, D. B., Sillanpää, M. J., & Savolainen, O. (2017). Genetic heterogeneity underlying variation in a locally adaptive clinal trait in *Pinus sylvestris* revealed by a Bayesian multipopulation analysis. *Heredity*, 118(5), 413–423. <https://doi.org/10.1038/hdy.2016.115>
- Kujala, S. T., & Savolainen, O. (2012). Sequence variation patterns along a latitudinal cline in Scots pine (*Pinus sylvestris*): Signs of clinal adaptation? *Tree Genetics and Genomes*, 8(6), 1451–1467. <https://doi.org/10.1007/s11295-012-0532-5>
- Kuparinen, A., Savolainen, O., & Schurr, F. M. (2010). Increased mortality can promote evolutionary adaptation of forest trees to climate change. *Forest Ecology and Management*, 259(5), 1003–1008. <https://doi.org/10.1016/j.foreco.2009.12.006>
- Laird, N. M., & Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. *Nature Reviews Genetics*, 7(5), 385–394. <https://doi.org/10.1038/nrg1839>
- Lande, R., & Porcher, E. (2017). Inbreeding depression maintained by recessive lethal mutations interacting with stabilizing selection on quantitative characters in a partially self-fertilizing population. *Evolution*, 71(5), 1191–1204. <https://doi.org/10.1111/evo.13225>
- Lande, R., Schemske, D. W., & Schultz, S. T. (1994). High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. *Evolution*, 48(4), 965–978.
- Langlet, O. (1971). Two hundred years genecology. *Taxon*, 20(5), 653–721. <https://doi.org/10.2307/1218596>
- Latta, R. G. (1998). Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *The American Naturalist*, 151(3), 283–292. <https://doi.org/10.1086/286119>
- Le Corre, V., & Kremer, A. (2003). Genetic variability at neutral markers, quantitative trait loci and trait. *Genetics*, 1219(July), 1205 – 1219.
- Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, 21(7), 1548 – 1566. <https://doi.org/10.1111/j.1365-294X.2012.05479.x>
- Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. (2014). Rare-variant association analysis: Study designs and statistical tests. *American Journal of Human Genetics*, 95(1), 5–23. <https://doi.org/10.1016/j.ajhg.2014.06.009>
- Lenz, P. R., Beaulieu, J., Mansfield, S. D., Clément, S., Desponts, M., & Bousquet, J. (2017). Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BMC Genomics*, 18, 335. <https://doi.org/10.1186/s12864-017-3715-5>
- Lewontin, R. C. (1974). *The genetic basis of evolutionary change* (Vol. 560). New York, NY: Columbia University Press.
- Li, D., Lewinger, J. P., Gauderman, W. J., Murcray, C. E., & Conti, D. (2011). Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genetic Epidemiology*, 35(8), 790–799. <https://doi.org/10.1002/gepi.20628>
- Li, Z., De La Torre, A. R., Sterck, L., Canovas, F.M., Avila, C., Merino, I., ... Van de Peer, Y. (2017). Single-Copy Genes as Molecular Markers for Phylogenomic Studies in Seed Plants. *Genome Biology and Evolution*, 9(5), 1130–1147. <https://doi.org/10.1093/gbe/evx070>
- Lim, K.-J., Paasela, T., Harju, A., Venäläinen, M., Paulin, L., Auvinen, P., ... Teeri, T. H. (2016). Developmental changes in Scots pine transcriptome during heartwood formation. *Plant Physiology*, 172(3), 1403–1417. <https://doi.org/10.1104/pp.16.01082>
- Lindén, A., Heinonsalo, J., Buchmann, N., Oinonen, M., Sonninen, E., Hiltavuori, E., & Pumpanen, J. (2014). Contrasting effects of increased carbon input on boreal SOM decomposition with and without presence of living root system of *Pinus sylvestris* L. *Plant and Soil*, 377(1–2), 145–158. <https://doi.org/10.1007/s11104-013-1987-3>
- Liu, H., Meuwissen, T. H. E., Sørensen, A. C., & Berg, P. (2015). Upweighting rare favourable alleles increases long-term genetic gain in genomic selection programs. *Genetics Selection Evolution*, 47(1), 19. <https://doi.org/10.1186/s12711-015-0101-0>
- Lönnroth, E. (1926). Untersuchungen über die innere Struktur und Entwicklung gleichaltriger naturnormaler Kiefernbestände: Basiert auf Material aus der Südhälfte Finnlands. *Acta Forestalia Fennica*, 30, 1–269.
- Lotterhos, K. E., Yeaman, S., Degner, J., Aitken, S., & Hodgins, K. A. (2018). Modularity of genes involved in local adaptation to climate despite physical linkage. *Genome Biology*, 19(1), 1–24. <https://doi.org/10.1186/s13059-018-1545-7>
- Lu, M., Krutovsky, K. V., Nelson, C. D., Koralewski, T. E., Byram, T. D., & Loopstra, C. A. (2016). Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.). *BMC Genomics*, 17(1), 1–11. <https://doi.org/10.1186/s12864-016-3081-8>
- Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits* (Vol. 1). Sunderland, MA: Sinauer.
- Ma, X. F., Hall, D., St. Onge, K. R., Jansson, S., & Ingvarsson, P. K. (2010). Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. *Genetics*, 186(3), 1033–1044. <https://doi.org/10.1534/genetics.110.120873>
- Maher, K. A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D. A., ... Deal, R. B. (2018). Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *The Plant Cell*, 30(1), 15–36. <https://doi.org/10.1105/tpc.17.00581>
- Mahony, C. R., MacLachlan, I. R., Lind, B. M., Yoder, J. B., Wang, T., & Aitken, S. N. (2019). Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. *BioRxiv*, 568725.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., ... Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747–753. <https://doi.org/10.1038/nature08494>
- Mason, W. L., & Alía, R. (2000). Current and future status of scots pine (*Pinus sylvestris* L.) forests in Europe. *Forest Systems*, 9(S1), 317–335.
- Mather, K. (1943). Polygenic inheritance and natural selection. *Biological Reviews*, 18(1), 32–64. <https://doi.org/10.1111/j.1469-185X.1943.tb00287.x>
- Mathieson, I., & McVean, G. (2012). Differential confounding of rare and common variants in spatially structured populations. *Nature Genetics*, 44(3), 243–246. <https://doi.org/10.1038/ng.1074>
- Matías, L., & Jump, A. S. (2012). Interactions between growth, demography and biotic interactions in determining species range limits in a warming world: The case of *Pinus sylvestris*. *Forest Ecology and Management*, 282, 10–22. <https://doi.org/10.1016/j.foreco.2012.06.053>
- Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favorable gene. *Genetical Research*, 23(23), 23–35. <https://doi.org/10.1017/s0016672300014634>
- McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, 17(4), 656–669. <https://doi.org/10.1111/1755-0998.12613>
- Mei, W., Stetter, M. G., Gates, D. J., Stitzer, M. C., & Ross-Ibarra, J. (2018). Adaptation in plant genomes: Bigger is different. *American Journal of Botany*, 105(1), 16–19. <https://doi.org/10.1002/ajb2.1002>

- Merino, I., Abrahamsson, M., Sterck, L., Craven-Bartle, B., Canovas, F., & von Arnold, S. (2016). Transcript profiling for early stages during embryo development in Scots pine. *BMC Plant Biology*, 16(1), 1–18. <https://doi.org/10.1186/s12870-016-0939-5>
- Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, 28(11), 659–669. <https://doi.org/10.1016/j.tree.2013.08.003>
- Meuwissen, T., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819–1829.
- Meuwissen, T., Hayes, B., & Goddard, M. (2016). Genomic selection – A paradigm shift in animal breeding. *Animal Frontiers*, 6, 6–14. <https://doi.org/10.2527/af.2016-0002>
- Mikola, J. (1982). Bud-set phenology as an indicator of climatic adaptation of Scots pine in Finland. *Silva Fennica*, 16, 178–184.
- Molotkov, P. I., & Patlaj, I. N. (1991). Systematic position within the genus *Pinus* and interspecific taxonomy. In M. Giertych, & C. Matyas (Eds.), *Genetics of Scots Pine* (pp. 31–40). Budapest, Hungary: Elsevier.
- Morgenstern, K. E. (1996). *Geographic variation in forest trees: Genetic basis and application of knowledge in silviculture*. Vancouver, Canada: UBC Press.
- Moritsuka, E., Hisataka, Y., Tamura, M., Uchiyama, K., Watanabe, A., Tsumura, Y., & Tachida, H. (2012). Extended linkage disequilibrium in noncoding regions in a conifer, *Cryptomeria japonica*. *Genetics*, 190, 1145–1148. <https://doi.org/10.1534/genetics.111.136697>
- Mosca, E., Eckert, A. J., Liechty, J. D., Wegrzyn, J. L., La Porta, N., Vendramin, G. G., & Neale, D. B. (2012). Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evolutionary Applications*, 5(7), 762–775. <https://doi.org/10.1111/j.1752-4571.2012.00256.x>
- Mullin, T. J., Andersson, B., Bastien, J. C., Beaulieu, J., Burdon, R. D., Dvorak, W. S., ... McKeand, S. E. (2011). Economic importance, breeding objectives and achievements. In I. C. Plomion, J. Bousquet, & C. Kole (Eds.), *Genetics, genomics and breeding of conifers*, (pp. 40–127). New York, NY: Edenbridge Science Publishers and CRC Press.
- Muona, O., & Harju, A. M. (1989). Effective population sizes, genetic variability, and mating system in natural stands and seed orchards of *Pinus sylvestris*. *Silvae Genetica*, 38, 221–228.
- Muona, O., & Szmidt, A. E. (1985). A multilocus study of natural populations of *Pinus sylvestris*. In H.-R. Gregorius (Ed.), *Lecture Notes in Biomathematics* (pp. 226–240). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Muona, O., Yazdani, R., & Rudin, D. (1987). Genetic change between life stages in *Pinus sylvestris*: Allozyme variation in seeds and planted seedlings. *Silvae Genetica*, 36, 39–42.
- Naydenov, K., Senneville, S., Beaulieu, J., Tremblay, F., & Bousquet, J. (2007). Glacial vicariance in Eurasia: Mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. *BMC Evolutionary Biology*, 7(1), 1–12. <https://doi.org/10.1186/1471-2148-7-233>
- Neale, D. B., & Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends in Plant Science*, 9(7), 325–330. <https://doi.org/10.1016/j.tplants.2004.05.006>
- Neale, D. B., Wegrzyn, J. L., Stevens, K. A., Zimin, A. V., Puiu, D., Crepeau, M. W., ... Langley, C. H. (2014). Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology*, 15(3), R59. <https://doi.org/10.1186/gb-2014-15-3-r59>
- Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Eileen Dolan, M., & Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genetics*, 6(4), e1000888. <https://doi.org/10.1371/journal.pgen.1000888>
- Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., ... Bergelson, J. (2005). The pattern of polymorphism in *Arabidopsis thaliana*. *PLOS Biology*, 3(7), e196. <https://doi.org/10.1371/journal.pbio.0030196>
- Notivol, E., García-Gil, M. R., Alia, R., & Savolainen, O. (2007). Genetic variation of growth rhythm traits in the limits of a latitudinal cline in Scots pine. *Canadian Journal of Forest Research*, 37, 540–551. <https://doi.org/10.1139/X06-243>
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., ... Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497(7451), 579–584. <https://doi.org/10.1038/nature12211>
- Ojeda Alayon, D. I., Mattila, T. M., Ruttink, T., Kujala, S. T., Kärkkäinen, K., Verta, J.-P., & Pyhäjärvi, T. (2018). Utilization of tissue ploidy level variation in *de novo* transcriptome assembly of *Pinus sylvestris*. *BioRxiv*. 95689. <https://doi.org/10.1101/495689>
- Oleksyn, J., Reich, P. B., Zytkowski, R., Karolewski, P., & Tjoelker, M. G. (2003). Nutrient conservation increases with latitude of origin in European *Pinus sylvestris* populations. *Oecologia*, 136(2), 220–235. <https://doi.org/10.1007/s00442-003-1265-9>
- Oleksyn, J., Zytkowski, R., Karolewski, P., Reich, P. B., & Tjoelker, M. G. (2000). Genetic and environmental control of seasonal carbohydrate dynamics in trees of diverse *Pinus sylvestris* populations. *Tree Physiology*, 20(12), 837–847. <https://doi.org/10.1093/treephys/20.12.837>
- Olson, M. S., Levsen, N., Soolanayakanahally, R. Y., Guy, R. D., Schroeder, W. R., Keller, S. R., & Tiffin, P. (2013). The adaptive potential of *Populus balsamifera* L. to phenology requirements in a warmer global climate. *Molecular Ecology*, 22(5), 1214–1230.
- Palme, A. E., Pyhäjärvi, T., Wachowiak, W., & Savolainen, O. (2009). Selection on nuclear genes in a *Pinus* phylogeny. *Molecular Biology and Evolution*, 26(4), 893–905. <https://doi.org/10.1093/molbev/msp010>
- Palmé, A. E., Wright, M., & Savolainen, O. (2008). Patterns of divergence among conifer ESTs and polymorphism in *Pinus sylvestris* identify putative selective sweeps. *Molecular Biology and Evolution*, 25(12), 2567–2577. <https://doi.org/10.1093/molbev/msn194>
- Palmroth, S., Berninger, F., Nikinmaa, E., Lloyd, J., Pulkkinen, P., & Hari, P. (1999). Structural adaptation rather than water conservation was observed in Scots pine over a range of wet to dry climates. *Oecologia*, 121(3), 302–309. <https://doi.org/10.1007/s004420050932>
- Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., ... Hayes, D., (2011). A large and persistent carbon sink in the world's forests. *Science*, 333, 988–993. <https://doi.org/10.1126/science.1201609>
- Parks, M., Cronn, R., & Liston, A. (2012). Separating the wheat from the chaff: mitigating the effects of noise in a plastome phylogenomic data set from *Pinus* L. (Pinaceae). *BMC Evol. Biol.*, 12(1), 100.
- Paulose, J., Hermisson, J., & Hallatschek, O. (2019). Spatial soft sweeps: Patterns of adaptation in populations with long-range dispersal. *PLoS Genetics*, 15(2), e1007936.
- Pavy, N., Namroud, M. C., Gagnon, F., Isabel, N., & Bousquet, J. (2012). The heterogeneous levels of linkage disequilibrium in white spruce genes and comparative analysis with other conifers. *Heredity*, 108(3), 273–284. <https://doi.org/10.1038/hdy.2011.72>
- Persson, B. (1998). Will climate change affect the optimal choice of *Pinus sylvestris* provenances? *Silva Fennica*, 32(2), 121–128. <https://doi.org/10.14214/sf.690>
- Persson, B., & Beuker, E. (1997). Distinguishing between the effects of changes in temperature and light climate using provenance trials with *Pinus sylvestris* in Sweden. *Canadian Journal of Forest Research*, 27(4), 572–579. <https://doi.org/10.1139/cjfr-27-4-572>
- Pool, J. E., & Aquadro, C. F. (2007). The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. *Molecular Ecology*, 16, 2844–2851. <https://doi.org/10.1111/j.1365-294X.2007.03324.x>
- Praciak, A. (2013). *The CABI encyclopedia of forest trees*. Oxfordshire, UK: CABI.
- Pravdin, L. F. (1969). *Scots pine: Variation, intraspecific taxonomy and selection*. London, UK: Ann Arbor Humphrey Science Publishers Ltd.
- Proost, S., Van Bel, M., Vanechoutte, D., Van De Peer, Y., Inze, D., Mueller-Roeber, B., & Vandepoele, K. (2015). PLAZA 3.0: An access

- point for plant comparative genomics. *Nucleic Acids Research*, 43(D1), D974–D981. <https://doi.org/10.1093/nar/gku986>
- Prunier, J., Pelgas, B., Gagnon, F., Despoints, M., Isabel, N., Beaulieu, J., & Bousquet, J. (2013). The genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. *BMC Genomics*, 14(1), 368. <https://doi.org/10.1186/1471-2164-14-368>
- Prus-Głowacki, W., Urbaniak, L., Bujas, E., & Curtu, A. L. (2012). Genetic variation of isolated and peripheral populations of *Pinus sylvestris* (L.) from glacial refugia. *Flora: Morphology, Distribution, Functional Ecology of Plants*, 207(2), 150–158. <https://doi.org/10.1016/j.flora.2011.11.006>
- Pyhäjärvi, T., García-Gil, M. R., Knürr, T., Mikkonen, M., Wachowiak, W., & Savolainen, O. (2007). Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics*, 177(3), 1713–1724. <https://doi.org/10.1534/genetics.107.077099>
- Pyhäjärvi, T., Kujala, S. T., & Savolainen, O. (2011). Revisiting protein heterozygosity in plants-nucleotide diversity in allozyme coding genes of conifer *Pinus sylvestris*. *Tree Genetics and Genomes*, 7(2), 385–397. <https://doi.org/10.1007/s11295-010-0340-8>
- Pyhäjärvi, T., Salmela, M. J., & Savolainen, O. (2008). Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. *Tree Genetics and Genomes*, 4(2), 247–254. <https://doi.org/10.1007/s11295-007-0105-1>
- Rehfeldt, G. E., Tchebakova, N. M., Parfenova, Y. I., Wykoff, W. R., Kuzmina, N., & Milyutin, L. I. (2002). Intraspecific responses to climate in *Pinus sylvestris*. *Global Change Biology*, 8(9), 912–929. <https://doi.org/10.1046/j.1365-2486.2002.00516.x>
- Reich, P. B., & Oleksyn, J. (2008). Climate warming will reduce growth and survival of Scots pine except in the far north. *Ecology Letters*, 11(6), 588–597. <https://doi.org/10.1111/j.1461-0248.2008.01172.x>
- Reich, P. B., Oleksyn, J., & Tjoelker, M. G. (1994). Seed mass effects on germination and growth of diverse European Scots pine populations. *Canadian Journal of Forest Research*, 24, 306–320. <https://doi.org/10.1139/x94-044>
- Rellstab, C., Dauphin, B., Zoller, S., Brodbeck, S., & Gugerli, F. (2019). Using transcriptome sequencing and pooled exome capture to study local adaptation in the giga-genome of *Pinus cembra*. *Molecular Ecology Resources*, 19(2), 536–551. <https://doi.org/10.1111/1755-0998.12986>
- Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, 273(5281), 1516–1517. <https://doi.org/10.1126/science.273.5281.1516>
- Ritland, K. (1996). Inferring the genetic basis of inbreeding depression in Plants. *Genome*, 39, 1–8. <https://doi.org/10.1139/g96-001>
- Robledo-Arnuncio, J. J., & Gil, L. (2005). Patterns of pollen dispersal in a small population of *Pinus sylvestris* L. revealed by total-exclusion paternity analysis. *Heredity*, 94(1), 13–22. <https://doi.org/10.1038/sj.hdy.6800542>
- Ruby, J. L., & Wright, J. W. (1976). A revised classification of geographic varieties in Scots pine. *Silva Fennica*, 25(5–6), 169–175.
- Rudin, D., Muona, O., & Yazdani, R. (1986). Comparison of the mating system of *Pinus sylvestris* in natural stands and seed orchards. *Hereditas*, 104(1), 15–19. <https://doi.org/10.1111/j.1601-5223.1986.tb00512.x>
- Saikkonen, K., Taulavuori, K., Hyvönen, T., Gundel, P. E., Hamilton, C. E., Vänninen, I., ... Helander, M. (2012). Climate change-driven species' range shifts filtered by photoperiodism. *Nature Publishing Group*, 2(4), 239–242. <https://doi.org/10.1038/nclimate1460>
- Salmela, M. J., Cavers, S., Cottrell, J. E., Iason, G. R., & Ennos, R. A. (2013). Spring phenology shows genetic variation among and within populations in seedlings of Scots pine (*Pinus sylvestris* L.) in the Scottish Highlands. *Plant Ecology and Diversity*, 6(3–4), 523–536. <https://doi.org/10.1080/17550874.2013.795627>
- Sarvas, R. (1962). Investigations on the flowering and seed crop of *Pinus sylvestris*. *Communications Instituti Forestalis Fenniae*, 53(4), 1–198.
- Savolainen, O., Bokma, F., García-Gil, R., Komulainen, P., & Repo, T. (2004). Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of *Pinus sylvestris* to climatic changes. *Forest Ecology and Management*, 197(1–3), 79–89. <https://doi.org/10.1016/j.foreco.2004.05.006>
- Savolainen, O., Kärkkäinen, K., Harju, A., Nikkanen, T., & Rusanen, M. (1993). Fertility variation in *Pinus sylvestris*: A test of sexual allocation theory. *American Journal of Botany*, 80(9), 1016–1020.
- Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, 38(1), 595–619. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095646>
- Sazonovs, A., & Barrett, J. C. (2018). Rare-variant studies to complement genome-wide association studies. *Annual Review of Genomics and Human Genetics*, 19(1), 97–112. <https://doi.org/10.1146/annurev-ev-genom-083117-021641>
- Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., ... Reyer, C. P. O. (2017). Forest disturbances under climate change. *Nature Climate Change*, 7(6), 395–402. <https://doi.org/10.1038/nclimate3303>
- Semerci, A., Semerci, H., Çalişkan, B., Çiçek, N., Ekmekçi, Y., & Mencuccini, M. (2017). Morphological and physiological responses to drought stress of European provenances of Scots pine. *European Journal of Forest Research*, 136(1), 91–104. <https://doi.org/10.1007/s10342-016-1011-6>
- Shutyaev, A. M., & Giertych, M. (1998). Height growth variation in a comprehensive Eurasian provenance experiment of (*Pinus sylvestris* L.). *Silvae Genetica*, 46(6), 332–349.
- Shutyaev, A. M., & Giertych, M. (2000). Genetic subdivisions of the range of Scots pine (*Pinus sylvestris* L.). *Silvae Genetica*, 3, 137–151.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Sinclair, W. T., Morman, J. D., & Ennos, R. A. (1999). The postglacial history of Scots pine (*Pinus sylvestris* L.) in western Europe: Evidence from mitochondrial DNA variation. *Molecular Ecology*, 8(1), 83–88. <https://doi.org/10.1046/j.1365-294X.1999.00527.x>
- Soranzo, N., Alia, R., Provan, J., & Powell, W. (2000). Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. *Molecular Ecology*, 9(9), 1205–1211. <https://doi.org/10.1046/j.1365-294X.2000.00994.x>
- Sorensen, F. C. (1982). The roles of polyembryony and embryo viability in the genetic system of conifers. *Evolution*, 36(4), 725. <https://doi.org/10.2307/2407885>
- Staab, P. R., & Metzler, D. (2016). Coala: An R framework for coalescent simulation. *Bioinformatics*, 32(12), 1903–1904. <https://doi.org/10.1093/bioinformatics/btw098>
- Stevens, K. A., Wegrzyn, J. L., Zimin, A., Puiu, D., Crepeau, M., Cardeno, C., ... Langley, C. H. (2016). Sequence of the sugar pine megagenome. *Genetics*, 204(4), 1613–1626. <https://doi.org/10.1534/genetics.116.193227>
- Stival Sena, J., Giguère, I., Boyle, B., Rigault, P., Birol, I., Zuccolo, A., ... Mackay, J. (2014). Evolution of gene structure in the conifer *Picea glauca*: A comparative analysis of the impact of intron size. *BMC Plant Biology*, 14(1), 95. <https://doi.org/10.1186/1471-2229-14-95>
- Suren, H., Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Smets, P., Rieseberg, L. H., ... Holliday, J. A. (2016). Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources*, 16, 1136–1146. <https://doi.org/10.1111/1755-0998.12570>
- Svenning, J. C., Normand, S., & Kageyama, M. (2008). Glacial refugia of temperate trees in Europe: Insights from species

- distribution modelling. *Journal of Ecology*, 96(6), 1117–1127. <https://doi.org/10.1111/j.1365-2745.2008.01422.x>
- Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, 105, 437–460.
- Torimaru, T., Wang, X. R., Fries, A., Andersson, B., & Lindgren, D. (2009). Evaluation of pollen contamination in an advanced scots pine seed orchard. *Silvae Genetica*, 58(5–6), 262–269. <https://doi.org/10.1515/sg-2009-0033>
- Tyrmi, J. (2018). STAPLER: A simple tool for creating, managing and parallelizing common high-throughput sequencing workflows. *BioRxiv*, <https://doi.org/10.1101/445056>
- Tyrmi, J., Vuosku, J., Acosta, J. J., Li, Z., Sterck, L., Savolainen, O., & Pyhäjärvi, T. (2019). Genomics of clinal local adaptation in *Pinus sylvestris* under continuous environmental and spatial genetic setting. *BioRxiv*, <https://doi.org/10.1101/647412>
- Väisänen, H., Kellomäki, S., Oker-Blom, P., & Valtonen, E. (1989). Structural development of *Pinus sylvestris* stands with varying initial density: A preliminary model for quality of sawn timber as affected by silvicultural measures. *Scandinavian Journal of Forest Research*, 4(1–4), 223–238. <https://doi.org/10.1080/02827588909382560>
- Varho, V., Rautiainen, A., Peltonen, M., Niemi, J., & Ovaska, U. (2018). *Biopaths to carbon neutrality* (Vol. 3). Publications of the Ministry of Agriculture and Forestry.
- Varis, S., Pakkanen, A., Galofré, A., & Pulkkinen, P. (2009). The extent of south-north pollen transfer in finnish Scots pine. *Silva Fennica*, 43(5), 717–726. <https://doi.org/10.14214/sf.168>
- Voronova, A., Belevich, V., Jansons, A., & Rungis, D. (2014). Stress-induced transcriptional activation of retrotransposon-like sequences in the Scots pine (*Pinus sylvestris* L.) genome. *Tree Genetics and Genomes*, 10(4), 937–951. <https://doi.org/10.1007/s11295-014-0733-1>
- Voronova, A., Belevich, V., Korica, A., & Rungis, D. (2017). Retrotransposon distribution and copy number variation in gymnosperm genomes. *Tree Genetics and Genomes*, 13(4), 88. <https://doi.org/10.1007/s11295-017-1165-5>
- Wachowiak, W., Balk, P. A., & Savolainen, O. (2009). Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). *Tree Genetics and Genomes*, 5(1), 117–132. <https://doi.org/10.1007/s11295-008-0188-3>
- Wachowiak, W., & Prus-Głowacki, W. (2008). Hybridisation processes in sympatric populations of pines *Pinus sylvestris* L., *P. mugo* Turra and *P. uliginosa* Neumann. *Plant Systematics and Evolution*, 271(1–2), 29–40. <https://doi.org/10.1007/s00606-007-0609-z>
- Wachowiak, W., Salmela, M. J., Ennos, R. A., Iason, G., & Cavers, S. (2011). High genetic diversity at the extreme range edge: Nucleotide variation at nuclear loci in Scots pine (*Pinus sylvestris* L.) in Scotland. *Heredity*, 106(5), 775–787. <https://doi.org/10.1038/hdy.2010.118>
- Wachowiak, W., Trivedi, U., Perry, A., & Cavers, S. (2015). Comparative transcriptomics of a complex of four European pine species. *BMC Genomics*, 16(1), 234. <https://doi.org/10.1186/s12864-015-1401-z>
- Wainschein, P., Jain, D. P., Yengo, L., Zheng, Z., TOPMed Anthropometry Working Group Trans-Omics for Precision Medicine Consortium, Cupples, L. A., ... Visscher, P. M. (2019). Recovery of trait heritability from whole genome sequence data. *BioRxiv*, 1–23. <https://doi.org/10.1101/588020>
- Wang, J., Zhou, Z., Zhang, Z., Li, H., Liu, D. I., Zhang, Q., ... Zhang, Z. (2018). Expanding the BLUP alphabet for genomic prediction adaptable to the genetic architectures of complex traits. *Heredity*, 121(6), 648. <https://doi.org/10.1038/s41437-018-0075-0>
- Wegrzyn, J. L., Lee, J. M., Teare, B. R., & Neale, D. B. (2008). TreeGenes: A forest tree genome database. *International Journal of Plant Genomics*, 2008, 412875.
- Wegrzyn, J. L., Liechty, J. D., Stevens, K. A., Wu, L. S., Loopstra, C. A., Vasquez-Gross, H. A., ... Neale, D. B. (2014). Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics*, 196(3), 891–909. <https://doi.org/10.1534/genetics.113.159996>
- Wiedemann, E. (1930). Die Versuche über den Einfluss der Herkunft des Kiefern-samens. *Zeitschr. F. Forst- Und Jagdwesen*, 62.
- Williams, C. G. (2009). *Conifer Reproductive Biology*. Dordrecht, The Netherlands: Springer. <https://doi.org/10.1007/978-1-4020-9602-0>
- Williams, C. G., & Savolainen, O. (1996). Inbreeding depression in conifers: Implications for breeding strategy. *Forest Science*, 42(1), 102–117.
- Willyard, A., Syring, J., Gernandt, D. S., Liston, A., & Cronn, R. (2007). Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution*, 24(1), 90–101. <https://doi.org/10.1093/molbev/msl131>
- Wójkiewicz, B., Cavers, S., & Wachowiak, W. (2016). Current approaches and perspectives in population genetics of scots pine (*Pinus sylvestris* L.). *Forest Science*, 62(3), 343–354. <https://doi.org/10.5849/forsci.15-040>
- Wright, J., & Ira, W. (1962). Geographic Variation in Scotch Pine. *Silvae Genetica*, 12, 1–25.
- Wu, H. X., Hallingbäck, H. R., & Sánchez, L. (2016). Performance of seven tree breeding strategies under conditions of inbreeding depression. *G3: Genes, Genomes, Genetics*, 6(3), 529–540.
- Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics*, 89(1), 82–93. <https://doi.org/10.1016/j.ajhg.2011.05.029>
- Yazdani, R., Muona, O., Rudin, D., & Szmidi, A. E. (1985). Genetic structure of a *Pinus sylvestris* L. seed-tree stand and naturally regenerated understory. *Forest Science*, 31(2), 430–436. <https://doi.org/10.1093/forestscience/31.2.430>
- Yeaman, S. (2015). Local adaptation by alleles of small effect. *The American Naturalist*, 186(S1), S74–S89. <https://doi.org/10.1086/682405>
- Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., ... Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, 353(6306), 1431–1433. <https://doi.org/10.1126/science.aaf7812>
- Zadworny, M., McCormack, M. L., Mucha, J., Reich, P. B., & Oleksyn, J. (2016). Scots pine fine roots adjust along a 2000-km latitudinal climatic gradient. *New Phytologist*, 212(2), 389–399. <https://doi.org/10.1111/nph.14048>
- Zhang, M., Zhou, L., Bawa, R., Suren, H., & Holliday, J. A. (2016). Recombination rate variation, hitchhiking, and demographic history shape deleterious load in poplar. *Molecular Biology and Evolution*, 33(11), 2899–2910. <https://doi.org/10.1093/molbev/msw169>
- Zhang, Z., Erbe, M., He, J., Ober, U., Gao, N., Zhang, H., ... Li, J. (2015). Accuracy of whole-genome prediction using a genetic architecture-enhanced variance-covariance matrix. *G3: Genes, Genomes, Genetics*, 5(4), 615–627. <https://doi.org/10.1534/g3.114.016261>
- Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D.-J., & Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS ONE*, 5(9), e12648. <https://doi.org/10.1371/journal.pone.0012648>
- Zhou, L., Bawa, R., & Holliday, J. A. (2014). Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (*Populus trichocarpa*). *Molecular Ecology*, 23, 2486–2499. <https://doi.org/10.1111/mec.12752>

How to cite this article: Pyhäjärvi T, Kujala ST, Savolainen O. 275 years of forestry meets genomics in *Pinus sylvestris*. *Evol Appl*. 2020;13:11–30. <https://doi.org/10.1111/eva.12809>