Spruce Population Genomics

Jean Bousquet, Sébastien Gérardi, Guillaume de Lafontaine, Juan Pablo Jaramillo-Correa, Nathalie Pavy, Julien Prunier, Patrick Lenz, and Jean Beaulieu

Abstract Spruce (*Picea* spp.) species are the dominant component of the circumboreal forest and one of the most reforested species groups in the world. They have become a reference among conifers for fundamental and applied genomics research. This chapter reviews the compelling progress made in the field of spruce population genomics, from the supportive field trials established by tree breeders to the release of complete sequences of their cytoplasmic and nuclear genomes to most recent applications in forestry. Initial efforts focusing on sequencing the spruce gene space resulted in the development of extensive genomic resources such as expressed sequence tags libraries, gene and single nucleotide polymorphism catalogs, genotyping arrays, and high-resolution genetic maps. During the last decade, these resources allowed to gain insights into a variety of topics such as phylogeography and phylogeny, introgression and speciation processes, as well as association mapping. Thanks to the recent advent of high-throughput genotyping and sequencing technologies, population genomics data are now being produced at an exponential rate, which translates into new applications and opportunities in conservation genetics and spruce breeding, such as genomic prediction.

J. Bousquet (✉) · S. Gérardi · N. Pavy · J. Prunier · J. Beaulieu
Canada Research Chair in Forest Genomics, Centre for Forest Research and Institute for Systems and Integrative Biology, Université Laval, Québec, QC, Canada
e-mail: jean.bousquet@sbf.ulaval.ca

G. de Lafontaine
Canada Research Chair in Integrative Biology of Northern Flora, Université du Québec à Rimouski, Rimouski, QC, Canada

J. P. Jaramillo-Correa
Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico

P. Lenz
Natural Resources Canada, Canadian Wood Fibre Centre, Québec, QC, Canada

Canada Research Chair in Forest Genomics, Université Laval, Québec, QC, Canada

Om P. Rajora (ed.), *Population Genomics: Forest Trees*,
Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2021_96,

Keywords Association mapping · Conifers · Conservation of genetic resources · Genome-wide association studies · Genomic selection · Hybrid zones · Introgression · Nucleotide diversity · Phylogeography · *Picea*

1 Introduction

Spruces (*Picea* spp.) are coniferous evergreen trees exclusively found in the Northern Hemisphere and due to their extensive distribution, they bear considerable ecological importance. They are a major component of boreal forests but are also found in mountainous areas of temperate zones, as well as in high elevation forests of subtropical areas. Because spruce forests also have an important economic value related to recreational activities and commercial use for a wide variety of timber and non-timber products, spruces have become one of the most planted species groups around the world (Mullin et al. 2011).

Population genomics is a relatively new discipline in forestry and forest genetics that can contribute to preserve the many benefits and functions of forests, especially in the context of rapid climate change currently affecting spruce forests worldwide. By providing new insights into processes driving population evolution, it can lead to improved monitoring and management of natural and domesticated genetic resources. Fueled by the advent of high-throughput sequencing and genotyping technologies, much progress has been made during the last 10 years regarding the advancement of spruce population genomics.

In this chapter, we comprehensively review milestones accomplished to date, which make spruces a potent model for conifer population genomics and its applications. We recapitulate the current state of knowledge regarding the biology and evolutionary history of the genus, then describe spruce genomic resources currently available for population genomic studies and the main research findings in population genomics, including their applications in the fields of genetic resource conservation and tree breeding. Finally, we look forward to major challenges and opportunities arising, and how they relate to the advancement of the spruce population genomics agenda and its integration with other 'omics research.

2 The Genus *Picea*

Picea is a member of the family Pinaceae, together with other widespread genera such as *Pinus*, *Abies*, *Larix,* and *Tsuga*. The family is an important component of the order Pinales, also known as Coniferales, the conifers (Gernandt et al. 2011). With approximately 38 extant species (Table 1), *Picea* is the third most diversified genus of the Pinaceae after *Pinus* and *Abies*.

Spruces (*Picea* spp.) are generally the dominant genus of the circumboreal forest (Farjon and Filer 2013), where they usually form large and contiguous populations

Table 1 Nomenclature of *Picea* spp. and their geographical distributions

Species	Subspecific classification		Natural range	Provenance tests/ breeding programs
	Liu's (1982) subgenus	Farjon's (2010) section		
P. abies (L.) Karst.	*Picea*	*Picea*	Baltico-Nordic boreal zone, Europe	√
P. alcoquiana (Veitch *ex* Lindl.) Carr.	*Picea*	*Picea*	Japanese Archipelago	
P. asperata Mast.	*Picea*	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. aurantiaca Mast.	*Picea*	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. brachytyla (Franch.) Pritzel	*Omorika*	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. breweriana S. Wats.	*Omorika*	*Picea*	Western North America	
P. chihuahuana Martínez	*Omorika*	*Picea*	Mexico	
P. crassifolia Komarov	*Picea*	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. engelmannii Parry *ex* Engelm.	*Picea*	*Casicta*	Western North America	√[a]
P. farreri Page & Rushforth	–	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. glauca (Moench) Voss	*Picea*	*Picea*	North American boreal zone	√
P. glehnii (Fr. Schmidt) Mast.	*Omorika*	*Picea*	Japanese Archipelago	
P. jezoensis (Sieb. & Zucc.) Carr.	*Picea*	*Casicta*	Eastern Asia, Japanese Archipelago	
P. koraiensis Nakai	*Omorika*	*Picea*	Eastern Asia	
P. koyamae Shirasawa	*Omorika*	*Picea*	Japanese Archipelago	
P. likiangensis (Franch.) Pritz.	*Picea*	*Casicta*	Qinghai-Tibetan Plateau and adjacent regions	
P. linzhiensis Cheng & Fu	–	*Casicta*	Qinghai-Tibetan Plateau and adjacent regions	
P. mariana (Mill.) Britt., Sterns & Poggenb.	*Picea*	*Picea*	North American boreal zone	√
P. martinezii Patterson	–	*Picea*	Mexico	

(continued)

Table 1 (continued)

Species	Subspecific classification		Natural range	Provenance tests/ breeding programs
	Liu's (1982) subgenus	Farjon's (2010) section		
P. maximowiczii Regel *ex* Mast.	*Omorika*	*Picea*	Japanese Archipelago	
P. mexicana Martínez	*Picea*	–	Mexico	
P. meyeri Rehd. & Wils.	*Picea*	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. morrisonicola Hayata	*Omorika*	*Picea*	Southeastern China	
P. neoveitchii Mast.	*Picea*	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. obovata Ledeb.	*Picea*	*Picea*	Eastern, Western Asia	√
P. omorika (Pancic) Purkyne	*Omorika*	*Picea*	Europe	
P. orientalis (L.) link	*Omorika*	*Picea*	Middle East	
P. pungens Engelm.	*Picea*	*Casicta*	Western North America	√
P. purpurea Mast.	*Picea*	*Casicta*	Qinghai-Tibetan Plateau and adjacent regions	
P. retroflexa Mast.	–	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. rubens Sarg.	*Picea*	*Picea*	Eastern North America	√
P. schrenkiana Fisch. & Mey.	*Omorika*	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. sitchensis (Bong.) Carr.	*Picea*	*Casicta*	Western North America	√
P. smithiana (Wall.) Boiss.	*Omorika*	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. spinulosa (Griff.) Henry	*Omorika*	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	
P. torano (Sieb. *ex* Koch) Koehne	*Picea*	*Picea*	Japanese Archipelago	
P. wilsonnii Mast.	*Omorika*	*Picea*	Qinghai-Tibetan Plateau and adjacent regions	

[a]Mostly focusing on *P. glauca* × *P. engelmannii* (interior spruce)

across Canada, Northern Europe, and Russia. Despite their extensive geographical distribution in the boreal zone, species diversity is relatively low due to the harsh environmental conditions prevailing at these latitudes (Peh et al. 2015). Hence, most spruce species are rather found in mountainous temperate forests of North America and Eurasia, with China being considered as a hot spot because it contains almost half of the total number of spruce species (Farjon and Filer 2013). In the southernmost part of their range, spruce species are restricted to subtropical high-altitude zones of Mexico and Taiwan where they typically occur as small and fragmented populations. However, distribution ranges are dynamic because species are constantly tracking climatically suitable habitats (Davis and Shaw 2001). This is especially relevant at large time scales, as Quaternary climatic oscillations prompted recurrent latitudinal or altitudinal range shifts of large amplitude (Hewitt 2000). Hence, current spruce distributions in boreal and subtropical zones mirror their idiosyncratic demographic trajectories, with ongoing climate change promoting range expansion in northern species (e.g., Payette and Filion 1985), and range contraction in southern species (Wright 1955; Jaramillo-Correa et al. 2006; Bodare et al. 2013). From a population genomics perspective, understanding the historical dynamics of species distributions is highly relevant since it strongly influences their extant natural genetic diversity and structure (Jaramillo-Correa et al. 2009; de Lafontaine et al. 2018).

As for other taxa in the *Pinaceae* family, spruce trees are monoecious. Self-fertilizing is therefore possible, although deleterious effects on viable seeds production were reported (Franklin 1970). However, due to high outcrossing rates, open-pollinated mating system, and wind-dispersed pollen and seeds, high heterozygosity and high gene flow among populations are usually observed in species from the temperate-boreal forests with large natural ranges (e.g., Isabel et al. 1995; Jaramillo-Correa et al. 2001; Perry and Bousquet 2001; O'Connell et al. 2007). This allows for the maintenance of large effective population sizes with high levels of recombination, which, in turn, results in high levels of genetic diversity. Exception to this rule are encountered in endemic spruce species with more fragmented distributions, reduced gene flow, and increased inbreeding where reduced genetic diversity and higher population differentiation are usually observed (Ledig et al. 2000a, b, 2002; Jaramillo-Correa et al. 2006, 2015).

Reproductive isolation among spruce taxa is generally weak and introgressive hybridization is common throughout the genus natural range (Wright 1955; Perron and Bousquet 1997; Rajora and Dancik 2000; Bouillé et al. 2011). Consequently, species delimitation is often difficult and many sympatric or parapatric species are rather considered as "species complexes" that can involve several taxa (e.g., Perron and Bousquet 1997; Du et al. 2009). Because of such weak reproductive isolation, interfertility is also maintained among various geographically isolated species, such as between the allopatric *P. glauca* (white spruce) and *P. jezoensis* (Wright 1955), two taxa currently occurring in North America and Asia, respectively. At the same time, species found in distinct phylogenetic clades but in sympatry over large areas are usually well isolated reproductively, as is the case for the largely distributed

P. glauca and *P. mariana* (black spruce) across the North American boreal forest (Haselhorst et al. 2019).

Both molecular and morphological data indicate that the genus *Picea* is monophyletic (e.g., Wright 1955; Hart 1987; Farjon 2010; Leslie et al. 2012; Li et al. 2017). The origin of the genus would date back to the Early Cretaceous, corresponding to the early diversification of the Pinaceae (Savard et al. 1994; Gernandt et al. 2008; Lin et al. 2010; Leslie et al. 2012). Based on several lines of evidence, the crown age of extant *Picea* lineages could be traced back much later to the Oligocene/Miocene transition (~25 Mya) (Bouillé and Bousquet 2005; Lin et al. 2010; Leslie et al. 2012; Lockwood et al. 2013; Ran et al. 2015). According to these studies, most extant spruce lineages diverged by the end of the Miocene (5–10 Mya), indicating that extant species radiation was relatively recent and rapid on the geological time scale. The geographic origin of the genus is still debated but most cpDNA and nucDNA phylogenies resolve the North American and morphologically archaic *P. breweriana* as sister to extant *Picea* lineages (Ran et al. 2006, 2015; Bouillé et al. 2011; Leslie et al. 2012), thus supporting the hypothesis of a North American origin of the genus. On the other hand, phylogenies relying on mtDNA are more conflicting, pointing to a North American (Ran et al. 2015) or Asian (Lockwood et al. 2013) origin of the genus depending on the set of DNA markers considered.

While monophyly of the genus is generally accepted, the taxonomic status of various species remains ambiguous, especially in Asia (Wright 1955; Ledig et al. 2004; Farjon 2008, 2010). Phylogenetic relationships among taxa are still a matter of debate, depending on which genome they relied upon (Sigurgeirsson and Szmidt 1993; Ran et al. 2006, 2015; Bouillé et al. 2011; Lockwood et al. 2013; Zou et al. 2016; Sullivan et al. 2017), and phylogenies are not congruent with morphological data, which are affected by parallel and convergent evolution (Liu 1982; Farjon 1990; Weng and Jackson 2000; Bouillé et al. 2011; Jia et al. 2014).

Several reasons related to population genetic processes may explain the lack of congruence among the many *Picea* phylogenies published to date. As for many northern hemisphere taxa, spruce species experienced repeated historical contacts via migration, given the various intercontinental land bridges that connected North America, Asia, and Eurasia during glacial eras since the late Jurassic (LePage 2003). In addition, introgression via long-distance pollen dispersal (e.g., Comtois 1997; Abbott and Brochmann 2003; Alsos et al. 2007) likely contributed to the complex biogeography. Reticulate evolution, which is the horizontal exchange of genetic material between species, has played a major role in plant evolution (Okuyama et al. 2004). In spruces, recurrent interspecific genetic exchanges likely impeded speciation (but see Sun et al. 2018) while maintaining weak reproductive isolation (e.g., Wright 1955; Mikkola 1969; Gordon 1976) and thus, facilitating introgression (e.g., Perron and Bousquet 1997; Du et al. 2009; Tollefsrud et al. 2015). Not only recent introgression but ancient reticulation (Bouillé et al. 2011) likely contributed to blur phylogenetic relationships among spruces. This is because the chloroplast genome and half of the nuclear genome are paternally transmitted through pollen in spruces, which facilitates long-distance dispersal and reticulate evolution, compared to

maternally transmitted mtDNA which is more locally dispersed through seeds (Bouillé et al. 2011; Sun et al. 2018). Given the recent crown age of extant *Picea* lineages and following the rapid radiation of extant species, incomplete lineage sorting (i.e., the retention of ancestral polymorphisms) has also been invoked to explain the lack of morphological and genetic divergence among *Picea* lineages and taxa. This process would thus complicate the reconstruction of the genus history (Bouillé and Bousquet 2005; Chen et al. 2010), but could also facilitate the conduct of population genomic studies by allowing the use of multi-species genotyping tools based on shared polymorphisms among species (Pavy et al. 2013a).

3 Spruce Population Genomic Resources

3.1 Field Trials

Many of the current population genomics applications in spruces would not be possible without the extensive field resources gathered in common-garden experiments initiated by spruce breeders in the last century. For instance, provenance trials allow phenotypic assessment of trees from various populations in homogenous environments, which is a prerequisite for genotype–phenotype association investigations aiming to identify marker loci that co-segregate with key phenotypic traits of adaptive or economical value (see Sect. 4 for further details). Genomic selection, which aims to estimate or predict the genetic worth of trees for quantitative traits (see Sect. 7 for further details), is another application that requires access to phenotypic data collected in genetic field tests. Further research into the adaptive potential of spruce populations to climate change will also benefit from using population genomics in connection with the retrospective analysis of reactions of common-garden populations to biotic or abiotic stress, in order to better design genomic scans such as association mapping. This is illustrated by recent studies in white spruce where tree ring analysis of mature provenance trees was used to identify signatures of past reactions to drought, how these reaction signatures were found to be related to local adaptation and use of population genomic scans to decipher genes and favorable alleles involved into local adaptation (Housset et al. 2018; Depardieu et al. 2020, 2021).

In addition, provenance/progeny trials represent a convenient and essential resource to sample populations across much of species ranges. Such extensive population sampling is often required to conduct meaningful population genomic studies, such as genotype-environment association studies (see Sect. 5 for further details), or studies of large-scale geographical structure in order to guide genetic resources conservation and management (see Sects. 6 and 8 for further details). Field trials are also critical resources from a conservation standpoint, as they allow preserving a large proportion of natural genetic diversity, which can prove useful to breeding programs as well as re-insertion of natural genetic diversity into the landscape, should the need arise. Field trials including large numbers of provenances

and families have been established in various climatic conditions and countries, especially for *P. abies*, *P. glauca*, *P. glauca* × *engelmannii* (interior spruce), *P. mariana*, *P. rubens* (red spruce), and *P. sitchensis* (Sitka spruce) (Mullin et al. 2011). A large provenance trial in Russia has also been reported and analyzed for *P. abies* and *P. obovata* (Nakvasina et al. 2017).

3.2 Spruce Genomic Resources Available for Population Genomics

Similarly to field resources, genomic resources are at the core of most population genomics investigations and applications. Thanks to large efforts and sequencing technology developments over the past 15 years, *Picea* nuclear genomes have been deciphered using a variety of approaches, and genetic markers from transcripts and exome sequences have been largely reported. Spruce genomic resources available to date include a variety of data and derived tools, such as genome sequences and annotated gene catalogs, DNA markers catalogs and genotyping chips, linkage maps or gene expression data. In the following section, we will summarize the main genomic resources of direct interest for spruce population genomics research and applications.

Spruces possess 2n = 24 chromosomes and their nuclear genome of ~20 Gb is among the largest ones in plants (Murray et al. 2012). Recently, next-generation sequencing (NGS) technologies have been used to obtain draft genome sequence assemblies for Norway spruce and white spruce (Birol et al. 2013; Nystedt et al. 2013; De La Torre et al. 2014a; Warren et al. 2015). In an effort to reduce genome complexity, gene catalogs were produced early on based upon large-scale expressed sequence tag (EST) sequencing, in order to describe the effective transcriptome of *Picea* species. Large-scale EST sequencing data were then combined to full-length cDNA sequencing data and assembled in transcript sequences to produce a large gene catalog of ~28,000 cDNA clusters in white spruce, representing as many distinct genes (Rigault et al. 2011). The number of transcribed spruce genes, currently estimated at ~32,000, is in the range of that observed in angiosperm model plants (Rigault et al. 2011; Nystedt et al. 2013), and it is composed of many gene families, often very large, containing conifer-specific and spruce-specific duplicates (Guillet-Claude et al. 2004; Bedon et al. 2010; Pavy et al. 2012a; De La Torre et al. 2015; Warren et al. 2015; Stival Sena et al. 2018). For Norway spruce, gene models were inferred along with the release of the draft genome sequence (Nystedt et al. 2013), thus providing opportunities to develop annotated marker resources in this species as well.

In addition to the nuclear genome, spruces possess two cytoplasmic genomes, the chloroplast and mitochondrial genomes. The advent of NGS technologies soon allowed sequencing whole cytoplasmic genomes and to gain much insights into their architecture and annotation. The first spruce complete chloroplast genome was

published by Cronn et al. (2008) for Sitka spruce (revised by Coombe et al. 2016), and complete sequences are now available for *P. abies* (Nystedt et al. 2013), *P. glauca* (Jackman et al. 2016), *P. jezoensis* (Yang et al. 2016), *P. engelmannii* (Lin et al. 2019), *P. crassifolia* and *P. asperata* (Ouyang et al. 2019), *P. mariana* (Lo et al. 2020), and *P. schrenkiana* (Li et al. 2020). With a size of around 124 kb, spruce chloroplast genomes show high sequence similarity among taxa, high synteny, and high homogeneity in gene content (114 genes). Similarly, complete mitochondrial genomes are now available for *P. abies* (Nystedt et al. 2013; Sullivan et al. 2020), *P. glauca* (Jackman et al. 2016), and *P. sitchensis* (Jackman et al. 2020). Spruce mitochondrial genomes are noticeably larger (5–6 Mb) than their chloroplast counterpart. Although their structure is complex and not fully resolved, much similarity has been observed in protein-coding genes among taxa. Over the years, these genomes have been the source of a large number of highly useful maternally inherited mtDNA and paternally inherited cpDNA polymorphisms in various spruce species (Table 2), allowing to infer their phylogeny, phylogeography, zones of suture and contact and more generally, subspecific structures (see Sects. 2 and 6).

Although highly useful for various population genetics applications, the first-generations of molecular markers from spruce nuclear genomes (e.g., allozymes, randomly amplified polymorphic DNAs (RAPDs), simple sequence repeats (SSRs) also called microsatellites, and expressed sequence tag polymorphisms (ESTPs)) were diverse but relatively limited in number. Over the past 10 years, the use of EST (expressed sequence tags) sequencing, exome capture and sequencing, RNA sequencing (RNA-Seq), genotyping-by-sequencing (GbS), and the development of automated pipelines of detection and filtering of single nucleotide polymorphisms (SNPs) allowed the development of large annotated SNP resources in white spruce, black spruce and more recently, in Norway spruce, which are at the basis of population genomics applications at the nuclear DNA level. A first annotated SNP resource including ~12,000 SNPs detected in a set of ~6,500 expressed genes was constructed for white spruce (Pavy et al. 2006). Following the discovery of nearly 400,000 nonsingleton gene SNPs in white spruce, a catalog of ~213,000 high-confidence SNPs spanning ~13,500 genes and ~2,500 gene families was released by Pavy et al. (2013b). Similarly, for black spruce, a catalog of ~100,000 high-confidence nonsingleton SNPs encompassing nearly 15,000 genes was obtained after the detection of ~460,000 SNPs from exome capture and sequencing (Pavy et al. 2016). Applying a similar approach to Norway spruce, a restricted set of ~62,000 high-confidence nonsingleton SNPs found in ~13,500 genes could be delimited from an initial set of ~240,000 SNPs obtained after exome capture and sequencing (Azaiez et al. 2018).

Other vast sets of SNPs were also obtained for Norway spruce and interior spruce. However, some of these SNP sets obtained in the context of targeted GbS or resequencing experiments were not meant to be annotated and released as SNP resources in the public domain. SNP sets were developed for Norway spruce from RNA-Seq (Heer et al. 2016) and from exome capture and sequencing (Chen et al. 2018; Mukrimin et al. 2018; Baison et al. 2019). Three other SNP sets were obtained for interior spruce from exome capture and sequencing (Suren et al. 2016), or from

Table 2 Polymorphic mitochondrial and chloroplastic DNA regions in *Picea* spp.

Species	Polymorphic mtDNA regions	References	Polymorphic cpDNA regions and markers	References
P. abies	*nad1* intron b/c	Sperisen et al. (2001)	cpSSR-*Pt*26081; cpSSR-*Pt*63718; cpSSR-*Pt*71936	Vendramin et al. (2000)
	mh02; mh05; mh08; mh09; mh09'; mh10; mh27; mh33; mh33'; mh34; mh35; mh38; mh44; mh50	Jeandroz et al. (2002)	*trnT-trnF*	Tollefsrud et al. (2015)
P. alcoquiana	*nad1* intron b/c; *mh02*	Aizawa et al. (2008)	*trnC-trnD; trnD-trnT; trnK*	Aizawa et al. (2008)
P. asperata	*nad1* intron b/c; *nad5* intron 1	Du et al. (2009)	–	–
P. chihuahuana	*mh10*	Jaramillo-Correa et al. (2006)	cpSSR-*Pt*26081; cpSSR-*Pt*30204	Jaramillo-Correa et al. (2006)
P. crassifolia	*nad1* intron b/c; *nad5* intron 1	Meng et al. (2007)	*trnC-trnD*	Meng et al. (2007)
			trnL-trnF; trnS-trnG; rps5-trnS; ndhK/C	Du et al. (2009)
P. glauca	*nad1* intron b/c; *nad5* intron 1	Jaramillo-Correa et al. (2003)	*trnL-trnF; trnT-trnL; ndhK/C*	Anderson et al. (2006)
P. glehnii	*nad1* intron b/c	Aizawa et al. (2015)	–	–
P. jezoensis	*nad1* intron b/c; *mh02*	Aizawa et al. (2007)	*trnC-trnD; trnD-trnT; trnK*	Aizawa et al. (2007)
P. koraiensis	–	–	*trnL-trnF; trnS-trnG; rps5-trnS; ndhK/C*	Du et al. (2009)
P. likiangensis	*nad1* intron b/c; *nad5* intron 1	Du et al. (2011)	*trnL-trnF; trnS-trnG; ndhK/C*	Du et al. (2011)
P. mariana	*nad1* intron b/c; *nad5* intron 1; *SSU rRNA* V1	Jaramillo-Correa et al. (2003)	cpSSR-*Pt*26081; cpSSR-*Pt*30204; cpSSR-*Pt*63718; cpSSR-*Pt*71936	Gérardi et al. (2010)
	nad7 intron 1	Jaramillo-Correa et al. (2004)		
P. mexicana	–	–	cpSSR-*Pt*26081; cpSSR-*Pt*63718	Jaramillo-Correa et al. (2015)
P. meyeri	–	–	*trnL-trnF; trnS-trnG; rps5-trnS; ndhK/C*	Du et al. (2009)

(continued)

Table 2 (continued)

Species	Polymorphic mtDNA regions	References	Polymorphic cpDNA regions and markers	References
P. morrisonicola	*nad1* intron b/c; *nad5* intron 1	Zou et al. (2013)	–	–
P. neoveitchii	*nad1* intron b/c; *nad5* intron 1	Zou et al. (2013)	*trnL-trnF; trnS-trnG; ndhK/C*	Zou et al. (2013)
P. obovata	*nad1* intron b/c	Tollefsrud et al. (2015)	*trnL-trnF; trnS-trnG; rps5-trnS; ndhK/C*	Du et al. (2009)
			trnT-trnF	Tollefsrud et al. (2015)
P. omorika	*nad1* intron b/c	Aleksić and Geburek (2014)	cpSSR-*Pt*26081; cpSSR-*Pt*71936	Nasri et al. (2008)
P. pungens	*nad1* intron b/c	Jaramillo-Correa et al. (2003)	–	–
P. purpurea	*nad1* intron b/c	Du et al. (2011)	*trnL-trnF; trnS-trnG; ndhK/C*	Du et al. (2011)
P. rubens	*nad1* intron b/c; *nad5* intron 1; *SSU rRNA* V1	Jaramillo-Correa et al. (2003)	cpSSR-*Pt*26081; cpSSR-*Pt*30204; cpSSR-*Pt*63718; cpSSR-*Pt*71936	Gérardi et al. (2010)
	nad7 intron 1	Jaramillo-Correa and Bousquet (2005)		
P. schrenkiana	*nad1* intron b/c; *nad5* intron 1	Li et al. (2015)	–	–
P. smithiana	*nad1* intron b/c; *nad5* intron 1	Li et al. (2015)	*trnL-trnF; trnS-trnG; ndhK/C*	Li et al. (2015)
P. wilsonii	*nad1* intron b/c; *nad5* intron 1	Zou et al. (2013)	*trnL-trnF; trnS-trnG; ndhK/C*	Zou et al. (2013)

sequencing barcoded genomic DNA (Chen et al. 2013; Gamal El-Dien et al. 2015). The four Norway spruce sets contained around 400,000 SNPs (Heer et al. 2016; Mukrimin et al. 2018) or more than 100,000 SNPs (Chen et al. 2018; Baison et al. 2019), but their annotation has remained limited. The interior spruce set obtained from exome capture and sequencing originally contained ~9 M SNPs, among which ~900,000 were deemed valid after applying several quality filtering criteria (Yeaman et al. 2016). The two remaining sets of interior spruce SNPs obtained after sequencing barcoded genomic DNA contained in excess of 1.2 M SNPs of unknown genomic location for the first set (~9,000 and ~63,000 deemed valid, depending on the imputation method used; Gamal El-Dien et al. 2015), and 1–1.5 M for the second one, among which ~18,000 SNPs had little or no missing data (Chen et al. 2013).

The availability of high-confidence annotated SNP resources in several species provided opportunities to further build reliable genotyping chips that could be used simultaneously for diverse population genomics applications (Table 3). Customized highly-multiplexed SNP genotyping arrays have first been developed for white spruce and black spruce (Pavy et al. 2008) as a mean for efficient, cost-effective, and repeatable genotyping of various natural and breeding populations for annotated gene SNPs. These genotyping chips came in different sizes based upon improving technologies and specific genotyping needs (Pavy et al. 2013a, 2016). Additional chips were developed subsequently for Norway spruce (Chen et al. 2012; Lind et al. 2014; Heer et al. 2016; Azaiez et al. 2018; Bernhardsson et al. 2021), as well as for Sitka spruce (Holliday et al. 2010b), for *P. sitchensis* × *P. glauca* (Hamilton et al. 2013a, b), and for interior spruce (De La Torre et al. 2014b, c), sometimes by using shared SNPs between these taxa and white spruce (Pavy et al. 2008, 2013a) so to hasten their development.

Genotyping arrays originally designed for white spruce were also tested on spruce species lacking adequate genomic resources (Pavy et al. 2013a), such as subtropical endemic species (Pavy et al. 2013a). The rate of SNP recovery between white spruce and seven other spruce taxa varied from 64 to 2% with a median of 17%, and it was positively influenced by decreasing phylogenetic distance and by the natural base level of molecular genetic diversity of the tested species (Pavy et al. 2013a). Bernhardsson et al. (2021) also reported that around half of the probes used to build a Norway spruce genotyping chip could also be called with high confidence in white spruce, black spruce, and Sitka spruce, which is a prerequisite for potential SNP detection with these probes.

In addition to their use in diverse population genomics studies (Table 3), the availability of high-quality annotated SNP resources and high-throughput (HT) genotyping arrays prompted the development of high-density linkage maps. Such maps are highly beneficial in many population genomic contexts, including studies focusing on the identification of quantitative trait loci (QTL) involved in the genetic control of various traits (e.g., in spruces, Pelgas et al. 2011, Lind et al. 2014, Pavy et al. 2017), on speciation and hybridization or on population genetic inferences (see Luikart et al. 2018 for a comprehensive review). Genetic mapping of spruce genomes was first conducted using mostly anonymous DNA markers (Tulsieram et al. 1992; Gosselin et al. 2002; Acheré et al. 2004; Scotti et al. 2005; Pelgas et al. 2006; Kang et al. 2010, 2011) and saturated genetic maps were soon obtained for white spruce, black spruce, and Norway spruce, recovering the expected 12 linkage groups. The era of HT transcriptome sequencing using RNA-Seq as well HT SNP genotyping technologies fostered a rapid expansion of genetic mapping in these three species, with the mapping of hundreds, then thousands of gene loci (Pavy et al. 2008, 2012a, 2017; Pelgas et al. 2011; Bernhardsson et al. 2019), and in some cases using large mapping populations of up to ~2,000 white spruce progeny so to attain high mapping precision (Pavy et al. 2017).

Table 3 High-throughput SNP genotyping arrays and resequencing methods used in spruce population genomics studies

Genotyping method	Species	Research topics	References
Sequenom iPLEX Gold arrays	*P. mariana*	Ecological genomics	Prunier et al. (2012)
	P. mariana	Association mapping	Prunier et al. (2013)
	P. mariana × *P. rubens*	Hybrid zones and introgression	de Lafontaine et al. (2015), de Lafontaine and Bousquet (2017)
	P. glauca	Pedigree traceability	Godbout et al. (2017)
Illumina GoldenGate arrays	*P. glauca*	Ecological genomics	Namroud et al. (2008)
	P. glauca, *P. mariana*	Linkage mapping	Pavy et al. (2008)
	P. glauca	Association mapping	Beaulieu et al. (2011)
	P. glauca	Linkage and QTL mapping	Pelgas et al. (2011)
	P. mariana	Ecological genomics	Prunier et al. (2011)
	P. abies	Ecological genomics	Chen et al. (2012)
	P. glauca	Conservation genetics	Namroud et al. (2012)
	P. glauca	Linkage mapping	Pavy et al. (2012a)
	P. sitchensis × *P. glauca*	Hybrid zones and introgression	Hamilton et al. (2013a, b)
	P. mariana	Association and QTL mapping	Prunier et al. (2013)
	P. glauca × *P. engelmannii*	Hybrid zones, and introgression	De La Torre et al. (2014a, b, 2015)
	P. abies	Linkage and QTL mapping	Lind et al. (2014)
Illumina Infinium iSelect arrays	*P. glauca* and 7 other *Picea* spp.	Multi-purpose	Pavy et al. (2013b)
	P. glauca	Genomic selection	Beaulieu et al. (2014a, b)
	P. glauca	Conservation genetics	Doerksen et al. (2014)
	P. glauca	Ecological genomics	Hornoy et al. (2015)
	P. abies	Ecological genomics	Heer et al. (2016)
	P. mariana	Validating exome SNP catalog	Pavy et al. (2016)
	P. mariana	Genomic selection	Lenz et al. (2017)
	P. glauca	Linkage and QTL mapping	Pavy et al. (2017)
	P. abies	Association mapping	Trujillo-Moya et al. (2018)
	P. abies	Validating exome SNP catalog	Azaiez et al. (2018)
	P. glauca	Genomic selection	Beaulieu et al. (2020)
	P. abies	Genomic selection	Lenz et al. (2020a)
	P. glauca	Genomic selection	Lenz et al. (2020b)

(continued)

Table 3 (continued)

Genotyping method	Species	Research topics	References
	P. mariana	Ecological genomics	Napier et al. (2020)
	P. glauca	Association mapping Ecological genomics	Depardieu et al. (2021)
	P. glauca	Seed orchard management	Galeano et al. (2021)
	P. glauca	QTL mapping	Laoué et al. (2021)
Affymetrix Axiom array	*P. abies* and 3 other *Picea* spp.	Multi-purpose	Bernhardsson et al. (2021)
Genotyping-by-Sequencing (GbS)	*P. glauca*	Multi-purpose	Chen et al. (2013)
	P. glauca × *P. engelmannii*	Genomic selection	Gamal El-Dien et al. (2015)
	P. glauca × *P. engelmannii*	Genomic selection	Ratcliffe et al. (2015)
	P. glauca × *P. engelmannii*	Multi-purpose	Suren et al. (2016)
	P. glauca × *P. engelmannii*	Association mapping Ecological genomics	Yeaman et al. (2016)
	P. abies	Genomic selection	Chen et al. (2018)
	P. abies	Association mapping	Mukrimin et al. (2018)
	6 Western North America *Picea* spp.	Phylogeography and phylogenomics	Haselhorst et al. (2019)
Exome capture and resequencing	*P. abies*	Association mapping	Baison et al. (2019)
	P. abies	Linkage mapping	Bernhardsson et al. (2019)
	P. abies	Phylogeography and phylogenomics	Chen et al. (2019a)
	P. abies	Association mapping Ecological genomics	Milesi et al. (2019)
	P. rubens	Phylogeography and phylogenomics	Capblancq et al. (2020)
	P. abies	Association mapping	Elfstrand et al. (2020)
	P. abies	Association mapping	Capador-Barreto et al. (2021)
RNA-Seq	*P. likiangensis*, *P. wilsonii*, *P. purpurea*	Phylogeography and phylogenomics	Ru et al. (2018)

4 Genotype–Phenotype Association Studies

Association mapping aims at deciphering genotype–phenotype associations (GPA) in unstructured or natural populations where linkage disequilibrium (LD) is usually very limited (Neale and Savolainen 2004; Pavy et al. 2012b). Using populations

replicated in common-garden studies, association mapping studies have been conducted under various settings in spruces, from testing modest numbers of candidate genes or candidate SNPs to testing large numbers of SNPs distributed over the whole genome, also known as genome-wide association studies (GWAS).

Most GPA studies in forest trees have tested the relationships between traits and genotypes using mixed linear regression models (Yu et al. 2006). They include the effect of the allele, the effect of identity-by-descent (expressed as a kinship matrix among individuals), and the effect of a potential hierarchical population structure (e.g., isolation by distance impeding random mating) (Yu et al. 2006; Holliday et al. 2010b; Beaulieu et al. 2011; Prunier et al. 2013; Sork et al. 2013; Trujillo-Moya et al. 2018). Corrections for multiple testing or false discovery rate (FDR) are usually applied to avoid false positives (Storey and Tibshirani 2003). Recursive partitioning algorithms such as Random Forest have also been used to identify restricted subsets of genetic polymorphisms that maximized the prediction of adaptive phenotypes by taking into account multiple individual loci effects, as well as interactions among them (Holliday et al. 2012).

GPA studies have been carried out for a variety of traits in spruces including those related to temperature adaptation such as growth, phenology, cold hardiness, and injuries (Holliday et al. 2010b; Prunier et al. 2013; Yeaman et al. 2016; Trujillo-Moya et al. 2018; Milesi et al. 2019; Depardieu et al. 2021) (Table 4). For instance, hundreds of SNPs from candidate genes were tested for relationships with adaptive traits in Sitka spruce (Holliday et al. 2010b) and in black spruce (Prunier et al. 2013). In both cases, a few dozens of genes were found associated with adaptive trait variation and showed allele frequency distributions correlated to latitudinal gradients, and overlap of gene annotations was found between sets of significant genes. As observed for most GPA studies in trees (Hall et al. 2016), individual SNPs explained a low proportion of the total trait variance in spruce species, typically from 1 to 5% for quantitative traits, with a higher range of values observed for simpler traits such as the concentrations of specific metabolites or biomolecules in spruce tissues (Ganthaler et al. 2017; Lamara et al. 2018) (Table 4).

GPA studies have also been conducted for drought-resistance traits in Norway spruce (Trujillo-Moya et al. 2018) and white spruce (Depardieu et al. 2021). Both studies reported a significant provenance effect on the phenotypic variation associated with drought response, which was shown to follow geographical and climatic clines (Trujillo-Moya et al. 2018; Depardieu et al. 2020). A number of gene SNPs were significantly associated with drought-related traits in white spruce, especially for those traits associated with post-stress recovery, including SNPs for genes shown to be differentially expressed in trees submitted to contrasted water-stress conditions (Depardieu et al. 2021).

GPA studies have also been conducted to identify SNPs associated with variation in wood traits. Beaulieu et al. (2011) conducted such a study by considering a large number of wood quality traits assessed in ~500 trees genotyped for ~500 candidate genes in white spruce. Only 13 genes were significantly associated with wood trait variations after false discovery rate (FDR) correction, each explaining between 1 and 5% of the total variation (Beaulieu et al. 2011). Lamara et al. (2016) tested about four

Table 4 Genotype–phenotype association studies in *Picea* spp.

Species	Traits	Nb. genets	Nb. genes screened	Nb. SNPs screened	Nb. significant SNPs (%)	PVE[a]	References
P. abies	Needle phenolic compounds	63	–	1,035	31 (3.0)	21.9–59.2	Ganthaler et al. (2017)
P. abies	Disease resistance	64	–	373,384	10	–	Mukrimin et al. (2018)
P. abies	Drought response, wood traits, climate-growth correlations	72	–	1,707	29	–	Trujillo-Moya et al. (2018)
P. abies	Wood traits	517	–	178,101	52 (0.03)	0.01–4.9	Baison et al. (2019)
P. abies	Growth, bud burst	777	–	917,107	387 (0.04)	–	Milesi et al. (2019)
P. abies	Fungal community on vegetative buds	478	–	178,101	9 (0.005)	–	Elfstrand et al. (2020)
P. abies	Pathogen resistance	330	–	63,760	18 (0.003)	0.05–0.06	Capador-Barreto et al. (2021)
P. glauca	Growth, wood traits	492	549	944	13 (1.4)	2.6–5.4	Beaulieu et al. (2011)
P. glauca	Wood traits	1,694	2,652	6,385	11 (0.2) – 401 (6.3)[b]	–	Lamara et al. (2016)
P. glauca	Needle acetophenones implicated in resistance to spruce budworm	211	2,312	4,747	35 (0.7)[c]	20–43[d]	Lamara et al. (2018)
P. glauca	Drought response, wood traits, climate-growth correlations	1,473	2,606	6,153	57 (0.9)	–	Depardieu et al. (2021)
P. glauca × *P. engelmannii*	Growth, budset, bud break, cold hardiness, dry mass	720	–	887,774	~500[e]	–	Yeaman et al. (2016)
P. mariana	Growth at several ages, budset	1,355	515	525	34 (66.7)[f]	0.1–2.9	Prunier et al. (2013)
P. sitchensis	Budset, cold hardiness	410	202	339	35 (10.3)[g]	0.7–5.4	Holliday et al. (2010b)
P. sitchensis	Budset, cold hardiness	410	202	339	20 (5.9)	30.1–37.2[h]	Holliday et al. (2012)

[a]Percent of variance explained
[b]Respectively using FDR correction, or no FDR correction for multiple testing in exploratory mode
[c]From multilocus association testing
[d]Single and multilocus estimates
[e]Approximate number of significant genome sequence contigs when correcting for population structure
[f]Small set of candidate genes and SNPs tested from previous identification from a large set of genes and SNPs using a regional genotype–environment association analysis and segregating sequence variants between DNA pools of contrasted phenotypes
[g]This study was in exploratory mode, implicating a less stringent statistical criterion to correct for multiple testing
[h]Cumulative variance explained by the top 20 SNPs using a Random Forest multilocus algorithm

times more white spruce trees and candidate gene loci for wood trait variation. Using more relaxed corrections for multiple testing, a large co-expression network could be identified, implicating key gene regulators. In Norway spruce, Baison et al. (2019) identified 39 significant genes by assessing genotype–phenotype associations for ~180,000 SNPs in more than 5,000 trees. In most of these studies, additional lines of evidence supporting the involvement of these genes in traits variation were obtained from gene expression experiments showing significant accumulations of transcripts in key tissues.

Other traits may be less continuously distributed than those analyzed above, which would indicate that they are controlled by a smaller number of genes with greater individual effects. Such a trend has been hypothesized for traits related to resistance to herbivorous insects and pathogens, where the production of a few compounds toxic to the pest can be sufficient to resist or sustain the attack. To verify this hypothesis, an association study was conducted in Norway spruce from Austria where trees harbored a binary resistant or non-resistant phenotype to needle bladder rust (*Chrysomyxa rhododendri*) (Ganthaler et al. 2017). A total of 20 SNPs were found significantly associated with rust resistance or variation in phenolic compounds, explaining, respectively, 22 and 59% of the phenotypic variance. The authors interpreted these results as evidence for narrower genetic control of rust resistance.

In a recent Norway spruce GWAS study relying on a clonal test, Mukrimin et al. (2018) identified eight genes (10 SNPs) involved in susceptibility to *Heterobasidion parviporum* infection, out of ~375,000 SNPs considered in the analysis (Table 4). The use of clonal replications permitted to obtain more precise phenotypic values. Three of the genes identified were previously reported as potentially involved in the regulation of plant defense responses, and hence appear as good candidates to further investigate susceptibility to fungal infection. While the study by Mukrimin et al. (2018) resulted in a very low rate of discovery, the reduced number of clones used for association analyses might have limited the statistical power, while enabling to genotype and test for much of the SNPs residing in the Norway spruce exome. Similar results were reported by Capador-Barreto et al. (2021) in a subsequent study on susceptibility to *Heterobasidion annosum* in the same species. Although no significant SNPs could be identified after using FDR correction, the use of a relaxed p-value threshold resulted in 18 significant SNPs, out of ~63,000 SNPs tested. While these significant loci were all different from those associated with resistance to *H. parviporum*, they identified 20 loci involved in resistance to both pathogens using a multi-trait GWAS approach.

Lamara et al. (2018) conducted a GPA study in white spruce for acetophenone metabolites, which are key defensive phenolic compounds related to chemical resistance against a major pest in eastern North America, the spruce budworm (*Choristoneura fumiferana* Clem.). They reported a total of eight significant single-locus gene associations for four different defense traits that explained from 2.3 to 11.2% of the phenotypic variance, depending on the metabolite. Using a multilocus approach, they could also identify 26 significant associations that explained up to 43% of the phenotypic variance, the best case being for the piceol

content of needles which implicated nine genes. Thus, the multilocus approach was more powerful for identifying candidate genes implicated in the constitutive defense against spruce budworm in white spruce. It also indicated that many genes with small to medium effects are likely implicated. Contrary to popular belief, no significant trade-offs with growth traits were noted.

Knowing that endophytic fungi in the phyllosphere have the potential to both enhance and reduce tree growth and fitness, Elfstrand et al. (2020) took another approach to study pathogen defense in Norway spruce. By combining barcoding of phyllosphere fungi with GWAS, they could identify nine markers associated with variations in fungal community and the presence of latent pathogens on dormant buds. Thus, the approach showed that these traits are determined not only by environmental factors, but also by host genetic factors.

5 Genotype–Environment Association Studies

Genotype–environment association (GEA) studies are an important part of ecological genomics and aim at identifying adaptive variation without prior information regarding the phenotypic traits under selection (Sork et al. 2013; Ćalić et al. 2016; Storfer et al. 2018). They include regressions, which test associations between allelic frequencies and environmental factors variation (e.g., Joost et al. 2007; Coop et al. 2010; Hornoy et al. 2015; Yeaman et al. 2016; Depardieu et al. 2021), recursive partitioning algorithms (e.g., Hornoy et al. 2015), or F_{ST}-based outlier detection analyses (Beaumont and Nichols 1996; Beaumont and Balding 2004) among others. The last approach requires a large number of genetic markers to model the neutral distribution of F_{ST} as a function of heterozygosity, and subsequently identify loci under selection departing from this neutral distribution.

Apart from the study of Bashalkhanov et al. (2013) reporting on the genetic response of red spruce populations to air pollution, most GEA studies in *Picea* spp. addressed the issue of local adaptation to climate (Table 5), with temperature and precipitation/aridity being the most intensively surveyed factors. This body of work provides important insights into the molecular mechanisms governing the evolution of adaptive evolution in the genus *Picea*. In addition, by mapping the current distribution of adaptive genetic diversity among populations, these studies are useful for spruce breeding and gene conservation applications in the face of ongoing climate change (Hornoy et al. 2015).

One of the first SNP outlier detection scans for undomesticated plant or animal species was conducted in white spruce (Namroud et al. 2008). In this early work, SNPs from ~350 candidate genes were used to investigate possible signatures of local adaptation in trees from six ecoregions mostly defined by temperature and precipitation variation. Using two F_{ST}-based outlier methods, the number of SNPs putatively under selection ranged between 0.9 and 9%, depending on the severity of statistical criteria used to identify outlier SNPs. These rates appear in line with values derived from similar genomic scans in other spruce species conducted since then

Table 5 Genotype–environment association studies in *Picea* spp.

Species	Research topics	Nb. samples	Nb. genes screened	Nb. SNPs screened	Nb. significant SNPs (%)	Methods	References
P. abies	Adaptation (latitudinal gradient)	303	232	445	29 (6.5)	Outlier detection + regression	Chen et al. (2012)
P. abies	Climate adaptation	846	224	237	43 (18.1)	Outlier detection + regression	Scalfi et al. (2014)
P. abies	Climate adaptation	826	172	214	13 (6.1)	Outlier detection + regression	Di Pierro et al. (2016)
P. abies	Adaptation (altitudinal gradient)	687	144	175	20 (11.4)	Outlier detection + regression	Di Pierro et al. (2017)
P. abies	Climate adaptation	777	–	917,107	835 (0.09)[a]	Regression	Milesi et al. (2019)
P. glauca	Adaptation (ecoregions)	158	345	534	2–49 (0.4–9.1)[b]	Outlier detection	Namroud et al. (2008)
P. glauca	Climate adaptation	198	7,819	11,085	43–144 (0.4–1.3)[b]	Random Forest + regression	Hornoy et al. (2015)
P. glauca	Drought and climate adaptation	1,473	2,606	6,153	437 (7.1)	RDA[c] + regression	Depardieu et al. (2021)
P. glauca × *P. engelmannii*	Climate adaptation	818	290	311	20 (6.4)	Outlier detection + regression	De La Torre et al. (2014b)
P. glauca × *P. engelmannii*	Climate adaptation	566	–	887,774	~300[d]	Outlier detection + regression	Yeaman et al. (2016)
P. mariana	Climate adaptation	156	313	583	26 (4.5)	Outlier detection + regression	Prunier et al. (2011)
P. mariana	Climate adaptation	593	463	473	23 (48.9)[e]	Outlier detection + regression	Prunier et al. (2012)
P. mariana	Drought adaptation	158	520	520	9 (1.7)	RDA[c] + regression	Napier et al. (2020)
P. rubens	Adaptation (air pollution)	48	36	61	7 (11.5)	Outlier detection + regression	Bashalkhanov et al. (2013)

[a]Number of unique transcripts
[b]Depending on severity of correction for FDR
[c]RDA, redundancy analysis
[d]Appoximate number of significant genome sequence contigs when correcting for population structure
[e]Reduced set of candidate genes and SNPs previously identified from a large set of genes and SNPs screened through a regional environmental association analysis and segregating sequence variants between DNA pools of phenotypic extremes related to growth and budset

(Table 5). Such differences seen in the discovery rate of putative adaptive SNPs from one study to the next can usually be attributed to the severity of correction for FDR, as observed in GPA studies. A notable exception is the study of Prunier et al. (2012) in black spruce, where a high rate of adaptive polymorphisms was reported (~50%), given that all gene SNPs tested had been already identified as outliers in a previous scan of a much larger set of SNPs conducted on a more restricted part of the black spruce transcontinental range (Prunier et al. 2011).

A larger GEA study was subsequently conducted in white spruce (Hornoy et al. 2015), implicating ~11,000 SNPs in ~8,000 expressed genes. Signatures of adaptation to climatic factors involved only a few dozen genes when severe FDR was employed, resulting in a very low rate of positive identification (~0.5%). Only small to moderate genetic effects and allele frequency shifts were observed, which suggested relatively weak selection intensities at individual loci, in line with previous per generation estimates (Prunier et al. 2011). However, taken together, these SNPs accounted for over 50% of the variance in climatic factors, but with more genes implicated in temperature adaptation than in adaptation to precipitation, which is a less contrasted climatic factor in Eastern North America where white spruce populations were sampled.

Similarly, a low rate of positive identification was obtained in an even larger GEA scan in interior spruce in British Columbia, where hundreds of genes carrying adaptive SNPs to climatic factors could be positively identified ~900,000 SNPs (Yeaman et al. 2016). Almost half of the putative adaptive genes previously identified in white spruce (Hornoy et al. 2015) could be found among these genes, indicating partly similar metabolic processes involved in climate adaptation in both species. To a fewer degree, some of the interior spruce genes could also be identified in the phylogenetically more remote but sympatric *Pinus contorta* (Yeaman et al. 2016), suggesting that genetic adaptation to climate could be constrained to some degree at the level of the family Pinaceae.

A very low rate of positive identification (in the 0.1% range) was also reported in a GEA study conducted in Norway spruce (Milesi et al. 2019), indicating that this is a common trend in widespread spruce taxa. In this large-scale study, the authors used ~900,000 SNPs to identify genetic variation involved in adaptation to temperature and precipitation in ~800 individuals sampled throughout the species range. More than 800 putatively adaptive polymorphisms were discovered after controlling for population structure and false positives. The overlap between markers showing significant associations with temperature and precipitation variables was limited, and contrary to white spruce, the majority of adaptive SNPs were related to precipitation variables. Altogether, approximately 50% of the markers were located in intergenic regions, 25% of them in intronic regions, while the remaining 25% was located in exons, which indicates that noncoding regions can have a significant role in spruce adaptive evolution.

In the face of climate change, drought has drawn increased attention in recent years, given that it has been identified as a major cause of decline in productivity and survival of forest species (Choat et al. 2012). Hence, drought has been studied in black spruce Alaskan populations, using over 500 gene SNPs by means of outlier

detection, redundancy analysis, and regression approaches (Napier et al. 2020). The intersect of the methods resulted in nine significant SNPs, representing as many distinct genes (see Sect. 5 for further details). Given that Alaska likely represents a distinct historical genetic lineage in black spruce (Gérardi et al. 2010), this result supports the idea that adaptation can build up from lineage-specific standing genetic variation (Prunier et al. 2012). Depardieu et al. (2021) also focused on drought adaptation in a recent white spruce population survey combining GEA, GPA, and transcriptomic approaches. Starting from a subset of ~6,500 gene SNPs, they identified 285 genes likely involved in drought resistance, including 110 genes differentially expressed in response to drought under greenhouse-controlled conditions. These genes represent valuable candidates for surveys in range-wide natural populations, as well as in breeding populations, given the scale of anticipated climate change in the next decades.

As for results from GPA studies, these trends suggest that genetic adaptation to climate is multi-dimensional and highly multigenic with small single-locus effects. These empirical results are coherent with theoretical expectations predicting small shifts in allele frequencies and little differentiation among populations (F_{ST}) at adaptive loci in species experimenting high gene flow and recent natural selection (Yeaman and Whitlock 2011; Le Corre and Kremer 2012; Rajora et al. 2016). These features typically apply to most spruce species due to wind pollination and large effective population sizes, and because most spruce species experienced recent range shifts and demographic fluctuations dating back to the end of the last ice age, which translates into a relatively small number of generations since recolonization of glaciated areas but sizeable natural selection intensity at outlier loci (Prunier et al. 2011).

Altogether, spruce genes found significantly associated with environmental variation cover a large array of molecular functions related to stress responses (biotic and abiotic), phenology, development/growth, or reproduction (e.g., Hornoy et al. 2015; Yeaman et al. 2016; Depardieu et al. 2021). Adaptive genes with no known homologs in angiosperms have also been commonly detected in spruces, suggesting distinct evolution of adaptive pathways.

GEA studies pose several challenges mainly related to spatial scale of analysis, environment heterogeneity, and population structure (Storfer et al. 2018). Indeed, a majority of studies are designed to survey shifts in allele frequency of populations sampled along a specific environmental gradient. For instance, misleading conclusions can be drawn if another environmental gradient overlaps the targeted gradient of interest. Solutions to overcome this drawback include the replication of the experimental design on two or more gradient locations or regions (Prunier et al. 2012; Di Pierro et al. 2017). In order to control for overlapping gradients and local adaptation, some authors (e.g., Prunier et al. 2011) have successfully applied a strategy consisting in grouping geographically remote populations occurring in similar environmental conditions. While avoiding redundancy, assessing numerous environmental variables representing various potential selective constraints should also alleviate this limitation.

Population structure can also bias the results from GEA studies by creating artificial gradients of non-adaptive nature. The origin of population structure can either be recent, most often indicating non-random mating among individuals (especially at local to regional scales), or be ancient, and therefore reflect historical biogeographic processes such as species range shifts or demographic fluctuations. In either case, assessing population structure prior to conducting GEA studies is mandatory in spruces, considering that most *Picea* taxa still carry the genetic imprint of historical processes from the Pleistocene and Holocene eras (see Sect. 6 for further details). In some cases, such population structure can follow environmental gradients like climatic clines, or it can implicate hybrid zones, which can make it difficult to disentangle adaptive variation from population structure (Chen et al. 2012; Yeaman et al. 2016). In such GEA studies, it also appears important to control for isolation by distance (IBD), in order to reduce the number of false positives (Hornoy et al. 2015). Several analytical methods now take into account neutral population structure as a covariate in the analytical framework to address this issue (e.g., Coop et al. 2010), though some have argued that such methods might result in high rejection rate of true positives in certain situations (e.g., Yeaman et al. 2016).

Several strategies and body of evidence are generally used to gain more robust insights from landscape and association genomics studies. The most common strategy consists in cross-validating the results obtained using different analytical methods. For instance, results of outlier detection methods, regression methods, or recursive partitioning algorithms are often compared with each other (e.g., Prunier et al. 2011, 2012; Scalfi et al. 2014; Hornoy et al. 2015; Di Pierro et al. 2017; Napier et al. 2020). Annotating candidate genes and performing gene ontology (GO) enrichment analysis represent other common validation procedures that strengthen the link between putative adaptive genes and selective pressures (e.g., Bashalkhanov et al. 2013; Hornoy et al. 2015; Di Pierro et al. 2016; Yeaman et al. 2016). Comparing lists of putative adaptive genes in different spruce taxa, such as those shared between *P. glauca* (Hornoy et al. 2015) and *P. glauca* × *engelmannii* (Yeaman et al. 2016), may also help identify key genes when very large numbers of candidate genes representing much of the transcriptome are tested. Similarly, comparing results from GPA and GEA analyses may help identify key genes given that large number of candidates are screened (Yeaman et al. 2016), although limited overlap is usually expected given the distinct physiological or metabolic processes, often largely unknown, that may underly testing in GPA versus GEA analyses (Depardieu et al. 2021). Also, expression studies at the genotypic level in controlled conditions represent one of the most convincing *a posteriori* demonstrations for the implication of specific gene polymorphisms in adaptation to environmental factors (Chen et al. 2012; Prunier et al. 2015; Yeaman et al. 2016) or in relation to specific adaptive traits, as recently shown by Depardieu et al. (2021) where four white spruce genes previously identified with both GPA and GEA approaches were also found differentially expressed between trees submitted to drought and non-drought conditions under greenhouse-controlled environment.

6 Molecular Genetic Variation of Historical Nature

Climatic oscillations of the Quaternary ice ages have caused recurrent reorganizations of biota, including large-scale displacements of taxa, changes in species composition, and altered biodiversity patterns from genes to the ecosystems (Huntley and Webb 1989; Webb and Bartlein 1992; Davis and Shaw 2001). The expansion of the ice sheets confined high-latitude species toward suitable environments (i.e., glacial refugia), followed by the recolonization of previously inhospitable territories when the ice sheets receded (Bennett et al. 1991). The lingering genetic imprint of range contractions and expansions (i.e., bottleneck effects and founder events) is still recognizable in most boreal and temperate taxa (e.g., Hewitt 2000; Taberlet et al. 1998; Jaramillo-Correa et al. 2009). Specifically, the last glacial period appears to be largely responsible for the current distribution of species wide scale neutral genetic diversity (Hewitt 2000; Petit et al. 2003). In such endeavors, molecular genetic signatures obtained from genetic variation in the cytoplasmic and nuclear genomes have been key to deciphering the population genetic history of spruce taxa.

6.1 Natural Hybridization, Introgression, and Speciation

One consequence of these recurrent glacial/interglacial range shifts is the possibility of secondary contact between previously allopatric species (Hewitt 2011). When interspecific reproductive barriers are permeable between two closely related taxa, natural hybridization and introgressive hybridization (i.e., introgression; repeated backcrossing of hybrids with parental species) can occur within such secondary contact zones (Rieseberg et al. 2007). In *Picea*, several taxa are found in parapatry with opportunity for genetic contact to occur in presence of weak reproductive isolation (Wright 1955). The application of molecular markers to the genetic characterization of contact zones between interbreeding spruce species has been initiated early to quantify introgression (e.g., Sutton et al. 1991; Krutovskii and Bergman 1995; Perron et al. 1995; Rajora and Dancik 2000). These hybrid zones thus represent natural laboratories for population genomics, offering the opportunity to study gene flow between species, and eventually providing useful insights on speciation and for orienting the conservation and management of natural genetic resources (Barton and Hewitt 1985; de Lafontaine et al. 2015; Taylor et al. 2015; Arnold 2016).

Gene exchange between closely related parapatric taxa ultimately results in gene capture, a process in which a gene from a donor species is transferred irreversibly into a recipient species. Similarly, cytoplasmic capture refers to the production of hybrids carrying the nuclear genome of a species and the cytoplasmic genome of another (Rieseberg and Soltis 1991). In *Picea*, mtDNA capture was reported in Asia, within the *Picea asperata* complex (*P. asperata*, *P. crassifolia*, *P. obovata*,

P. meyeri, *P. koraiensis*) (Du et al. 2009) and between *P. purpurea* and *P. likiangensis* (Du et al. 2011), also in boreal Eurasia between *P. obovata* and *P. abies* (Tsuda et al. 2016), in Japan between *P. jezoensis* and *P. glehnii* (Aizawa et al. 2015, 2016), in western North America between *P. glauca* and *P. sitchensis* (Hamilton and Aitken 2013), and in eastern North America between *P. mariana* and *P. rubens* (Jaramillo-Correa and Bousquet 2005; Gérardi et al. 2010). However, in most instances, species delimitation relied on morphological evidence rather than molecular inferences or signatures drawn from the nuclear genome, which represents an over-simplification of a complex reality. Current population genomics approaches could help overcome this limitation by providing an accurate assessment of the genome-wide ancestry of individuals, which may in turn provide new insights into this process.

P. mariana and *P. rubens* represent a recent progenitor-derivative species pair, where molecular genetic evidence has shown that red spruce originated from glaciation-induced isolation of a pre-existing black spruce population during the Pleistocene (Perron et al. 2000; Jaramillo-Correa and Bousquet 2003). Although well differentiated morphologically (Manley 1971; Perron et al. 1995; Johnsen et al. 1998), a high occurrence of hybrids and highly introgressed individuals was reported from populations located in the large sympatric area, indicating weak reproductive isolation between the two taxa (Perron and Bousquet 1997). This finding was recently corroborated by de Lafontaine and Bousquet (2017), who reported a hybridization rate of 38% for ~200 trees sampled within the sympatric zone, using SNPs from ~300 genes spread over the 12 spruce chromosomes. However, the level of introgression between these two spruce taxa appears variable across the genome (de Lafontaine et al. 2015). Specifically, while most SNPs (79%) reflected selectively neutral diffusion across the porous species barrier, a small number of SNPs (8%) pointed at genomic regions virtually impermeable to interspecific gene flow, thus forming the actual genomic boundary between species. A third set of SNPs (13%) were highly permeable and were equally found in the two taxa (de Lafontaine et al. 2015). Such heterogeneous patterns of introgression across the genome were also reported in western North America between *P. glauca* and *P. sitchensis* (Hamilton et al. 2013a), as well as between *P. glauca* and *P. engelmannii* (De La Torre et al. 2014b).

Asymmetric patterns of interspecific gene flow have been reported in many spruce species complexes. For instance, in western Canada, *P. sitchensis* x *P. glauca* hybrids showed a predominance of Sitka spruce ancestry (Hamilton et al. 2013a). A similar trend was observed in the hybrid zone between the temperate/boreal *P. glauca* and the subalpine *P. engelmannii* along the Canadian Rocky mountains, where hybrids collectively showed a greater genetic contribution from the latter species than from *P. glauca* (De La Torre et al. 2014c). However, this finding was not confirmed by a recent study that analyzed populations sampled across a wider geographical range, suggesting that the direction of gene flow is likely heterogeneous in this species complex (Haselhorst et al. 2019). On the Qinghai–Tibet Plateau, directional introgression was also observed among various taxa pairs of the *P. likiangensis* species complex (Ru et al. 2018; Sun et al. 2018).

In another population genomics study, de Lafontaine and Bousquet (2017) estimated the asymmetrical migration rates among black spruce natural purebreds, red spruce natural purebreds, and their hybrids. They reported a higher rate of gene flow from black spruce natural purebreds toward red spruce natural purebreds than vice versa. This asymmetry likely reflects the historical, unidirectional gene flow between the two taxa at the time of species inception and during postglacial colonization. By contrast, greater level of gene flow was found between red spruce natural purebreds and the hybrids than between black spruce natural purebreds and the hybrids, particularly at loci impermeable to introgression. This observation suggests that the actual boundary between species is unidirectional and mostly preventing introgression with black spruce (de Lafontaine and Bousquet 2017). While population genomics provides valuable insights into natural hybridization, introgression, and speciation processes, Haselhorst et al. (2019) emphasized the importance of sampling individuals throughout the whole interspecific zone of contact, as intraspecific variation can bias the assessment of introgression patterns.

The above findings provided useful insight into the maintenance of hybrid zones and species integrity in *Picea*. Associations between hybrid index and environmental variables are usually invoked as a line of evidence that exogenous selection, that is, natural selection acting along environmental gradients or clines, is likely the main mechanism promoting the maintenance of hybrid zones (De La Torre et al. 2014c). The rationale is that both parental species are locally adapted to their respective native ecological niche and hybrids have higher fitness than either parent taxa within the transitional environment. This pattern may be reinforced by the apparent lack of phenotypic effect of parental species deleterious alleles due to complementation of deleterious alleles and reduced mutation load in hybrids (Conte et al. 2017). This model could explain the maintenance of the hybrid zone and species integrity in hybridizing spruce species pairs from north-western North America (e.g., *P. sitchensis* × *P. glauca* (Hamilton et al. 2013b; Hamilton and Aitken 2013), and *P. glauca* × *P. engelmannii* (De La Torre et al. 2014c). In the *P. mariana* × *P. rubens* case study, an association between ancestry and precipitation was found, with red spruce natural purebreds receiving higher precipitation than hybrids and black spruce purebreds within the zone of sympatry (de Lafontaine and Bousquet 2017). Although this association could suggest that exogenous selection is at play in this system, this interpretation falls short in light of the asymmetrical levels of gene flow. Specifically, precipitation could not stand out as a strong selective filter because i) gene flow between red spruce natural purebreds and the hybrids was globally high despite significant differences in precipitation, whereas ii) the amount of precipitation received was not different between black spruce natural purebreds and hybrids, that is within the actual barrier to interspecific gene flow (de Lafontaine and Bousquet 2017). The study of the hybrid zone between black spruce and red spruce reiterates that, while genetic/climatic associations should indeed be expected in the presence of exogenous selection, they do not provide firm evidence for this evolutionary process. In fact, formal assessment would require testing the fitness of progeny from controlled crosses in multiple environments across the hybrid zone, which represents a significant endeavor.

6.2 Phylogeography and Historical Demography

Phylogeography is the science of retracing the various ancestral lineages making up the modern genetic heritage of species (Jaramillo-Correa et al. 2009). Such knowledge has important implications for our understanding of the evolution of genetic diversity, for better orienting gene conservation efforts, and for assisting the delineation of breeding zones. Most early phylogeographic inferences were drawn using DNA markers from the chloroplast and mitochondrial genomes, due to the uniparental inheritance and contrasted dispersal of the two cytoplasmic genomes in *Picea* spp. and other Pinaceae. Indeed, paternally inherited cpDNA allows tracking pollen dispersal while maternally inherited mtDNA allows tracking seed dispersal, which is usually on more restricted distances than pollen dispersal on average, thus allowing to better track phylogeographic patterns in *Picea* spp. (e.g., Du et al. 2009; Gérardi et al. 2010). Along with pollen and macrofossil records evidence, a number of broad-scale phylogeographic surveys made it possible to uncover patterns of glacial survival, postglacial recolonization, and demographic fluctuations for most spruce taxa, as well as key vicariance factors that likely affected the genetic structure of a large numbers of plant and tree species (e.g., Jaramillo-Correa et al. 2009). Below, we review a number of compelling cases on spruce taxa with broad distributions, and for which genomic information has shed new light on the history of spruce taxa.

In North America, cytoplasmic DNA provided evidence for the persistence of the transcontinental black spruce through the last glacial maximum (LGM) in at least three largely dispersed glacial refugia south of the ice sheet, and also in two possible cryptic northern refugium located in unglaciated coastal Labrador and Beringia (Jaramillo-Correa et al. 2004; Gérardi et al. 2010). The range-wide comparison of the geographic extent of cytoplasmic DNA lineages revealed that extensive pollen gene flow between ancestral lineages occurred preferentially from west to east during the postglacial expansion of black spruce, while seed-mediated gene flow remained geographically restricted. The survey of cpDNA variation in white spruce populations also provided support for a Beringian glacial refugia, as well as bidirectional mixing along the putative migration route (Anderson et al. 2006). Additional studies showed that central and Eastern North America were colonized by genetically differentiated lineages originating from two southern glacial refugia possibly located in the Mississippi River region and east of the Appalachians, respectively (de Lafontaine et al. 2010).

More recently using a GbS approach, Haselhorst et al. (2019) surveyed patterns of genetic diversity and gene flow among six North American spruce taxa (*P. breweriana, P. pungens, P. sitchensis, P. engelmanii, P. mariana,* and *P. glauca*). Interspecific gene flow was confirmed between *P. glauca* and *P engelmannii*, but contrary to previous multiple evidence, no gene flow was observed between *P. glauca* and *P. sitchensis*, possibly due to incomplete sampling of the *P. sitchensis* natural range. The authors also identified three genetically differentiated clusters in *P. engelmannii*, while only two clusters were reported in this species to date (Ledig et al. 2006), thus illustrating how phylogeographic inferences can be

improved by combining high-throughput sequencing of the nuclear genome and genomic analyses.

Recently, Capblancq et al. (2020) investigated the geographic structure and demographic history of *P. rubens* across its fragmented natural range. The authors used an exome capture/sequencing approach to identify and analyze over 1 M SNPs in ~350 individuals (although such large number of nuclear markers are not necessary to detect neutral population structure). Using a PCA method, they could identify three differentiated genetic clusters distributed latitudinally along the species' range, thus located at the core, margin, and edge of red spruce distribution. Estimates of genetic divergence among the three clusters supported the idea that the species survived the LGM in a single glacial refugium, thus ruling out the hypothesis of multiple refugia in the Appalachian Mountains. Consistent with previous literature and evidence that red spruce would have originated from a black spruce isolated coastal population during the Pleistocene era of repeated glaciations (Perron et al. 2000; Jaramillo-Correa and Bousquet 2003), a sustained population decline initiated at least 700 kya during the Pleistocene era was observed, along with a major bottleneck dating back to 400 kya during one of the previous interglacial periods (Capblancq et al. 2020). A similar bottleneck at about the same time was also inferred in the same species from cpDNA microsatellites (Jaramillo-Correa et al. 2015). However, effective population size of the core cluster appeared stable during the last 3,000–5,000 years, suggesting that these populations do not require particular conservation efforts, contrary to their more southern counterparts (Capblancq et al. 2020). It also appears likely that the northernmost cluster might be characterized by significant introgression from black spruce (Capblancq et al. 2020), which is in agreement with similar observations using more limited sets of nuclear markers (Perron and Bousquet 1997; de Lafontaine et al. 2015), and also with the observation of directional gene flow from black spruce purebreds to red spruce purebreds in the zone of contact (de Lafontaine and Bousquet (2017).

Norway spruce is widespread in Europe and its range is divided into a northern part covering Fennoscandia and European Russia, and a southern part mainly occurring along the mountain ranges of central and southeastern Europe. From multi-gene sequence signatures, various bottlenecks were inferred at different times of the Pleistocene for these populations (Heuertz et al. 2006; Namroud et al. 2010). Early works based on cytoplasmic variation and paleobotanical evidence indicated that the northern part of the range originates from a single glacial refugium likely situated in Russia (Tollefsrud et al. 2008). Based on molecular and ancient DNA evidence, it was also proposed that Norway spruce could have persisted during the LGM in a western Scandinavia cryptic refugium located north of the ice sheet (Parducci et al. 2012). The southern part of the range would have been colonized by two or more distinct refugia possibly located in the Alps and the Carpathian Mountains (Tollefsrud et al. 2008).

Chen et al. (2019a) revisited recently the population structure and demographic history of Norway spruce using ~400,000 SNPs detected in an exome capture/sequencing experiment. They also included samples from the hybridizing species *P. obovata* to search for signatures of past introgression events (i.e., shared ancestral

polymorphisms), as well as from *P. omorika,* in order to assess the impact of recent material transfer from mainland Europe to Sweden, that occurred as part of Swedish reforestation programs. Their results essentially corroborated the three differentiated genetic clusters described previously, but showed that admixture among clusters was much more prevalent than previously thought. They also found that historical admixture from *P. obovata* into *P. abies* extended as far west as in Swedish populations. Taken together, these findings challenge the idea that postglacial recolonization was mainly initiated from populations that persisted in southern refugia, and the hypothesis that sparse populations may have survived further north is emerging, as shown in North America for the transcontinental black spruce (Jaramillo-Correa et al. 2004; Gérardi et al. 2010) and white spruce (Anderson et al. 2006). In addition to phylogeographic insights, Chen et al. (2019a) could identify Norway spruces with foreign genetic background in southern Sweden and Denmark, suggesting that genomic profiles are well suited to trace back relatively recent translocation events.

Phylogeographic surveys relying on cytoplasmic markers have also been conducted in numerous Asian taxa, which occur in three distinct geographical zones, namely the Japan archipelago, southeastern Russia/northeastern China, and the Qinghai-Tibetan Plateau (QTP). Well-differentiated intraspecific lineages have been reported in most species (e.g., Aizawa et al. 2007, 2008), and these species were also affected by Pleistocene climate oscillations, as observed for taxa in other parts of the world. Many pairs of phylogenetically close taxa found in parapatry, with inland and coastal or island distribution patterns, could represent cases of recent species divergence with incomplete reproductive isolation (Bouillé et al. 2011). Phylogeographic patterns are particularly complex in the QTP of northern China, which is a biodiversity hotspot home of multiple spruce species often forming interbreeding species complexes (e.g., Du et al. 2009). In this context, population genomics approaches appear particularly well suited to decipher the intricate relationships among these recently diverged taxa.

One example of such application was provided by Ru et al. (2018), who sampled 34 populations from the three closely related species *P. likiangensis*, *P. purpurea*, and *P. wilsonii*, and used phylogenetic and coalescent analyses to resolve their history. They found out that *P. purpurea* emerged from homoploid hybrid speciation from *P. likiangensis* and *P. wilsonii*, and that the species is now well delimited genetically. They also showed that around 60 and 40% of *P. purpurea* ancestry derived from *P. wilsonii* and *P. likiangensis*, respectively, thus providing evidence for asymmetric introgression between the two parental species. In addition, they discovered that the speciation process likely unfolded in two steps, involving the early divergence of a "ghost" (i.e., now extinct) lineage genetically closer to *P. likiangensis*, which would have subsequently evolved in *P. purpurea* following backcrossing events with *P. wilsonii*.

Patterns of Holocene expansions and modern gene flow between congeneric boreal conifers are still a matter of debate. For instance, most studies suggest that the deglaciated territory was primarily colonized by an advancing front, thus generating clinal patterns of spatial genetic structure (e.g., Chen et al. 2010; Holliday et al.

2010a; Hornoy et al. 2015). However, the contribution of long dispersal founding events is also acknowledged although many phylogeographic studies in *Picea* spp. have failed to detect them. Support for such phenomenon from mtDNA signatures has been provided in the boreal black spruce (Gamache et al. 2003). More recent analyses combining dendrochronology with next-generation sequencing data have also been successful in detecting such long-distance colonization events and confirming that pollen-mediated gene flow promotes high levels of genetic variation from the very beginning in the colonizing stands (Holliday et al. 2017). Such gene flow would provide the necessary variation for rapid adaptation to new environments, which in turn may generate new clines of adaptive alleles (e.g., Chen et al. 2012; Prunier et al. 2012; Hornoy et al. 2015).

As for patterns observed in the collapsing mountain populations of subtropical *Picea* species in Mexico, Quiñones-Pérez et al. (2014) found significant IBD among the southernmost stands of the endangered *P. chihuahuana*. These results confirmed previous evidence from cytoplasmic DNA markers of reduced genetic diversity and strong genetic isolation of the scattered stands from collapse due to Holocene warming (Jaramillo-Correa et al. 2006). For the more largely distributed *P. mexicana*, bottlenecks also could be inferred as far back as the previous interglacial period, more than 100 kya (Jaramillo-Correa et al. 2015).

6.3 Nucleotide Diversity

Demographic fluctuations related to population expansions or contractions are often detected by analyzing nucleotide diversity in DNA sequence data or by using coalescence simulations. To date, nucleotide diversity in *Picea* spp. has been largely estimated by resequencing of nuclear coding genes (Table 6). As expected from neutral theory, a recent study by Eklöf et al. (2020) in Norway spruce showed that nucleotide diversity estimates derived from whole-genome resequencing (WGR), exome sequencing, and GbS were comparable, indicating some interesting comparisons among spruce species. For instance, the nucleotide diversity of genes is usually lower in species secluded to mountain populations, such as *P. scherenkiana* or *P. breweriana*, and higher in geographically widespread species, such as white spruce or Norway spruce (e.g., Li et al. 2015; Conte et al. 2017; see also Table 6). However, some exceptions exist, particularly for species that have experienced drastic changes in their effective population sizes in the past. One such example is Sitka spruce, a species that spans along the Pacific coast of North America from northern California to Alaska, for which a strong historical bottleneck followed by a recent expansion was inferred from over 150 genes (Holliday et al. 2010a; Haselhorst et al. 2019). On the other hand, an unexpectedly high nucleotide diversity was estimated for *P. purpurea*, fueled since the middle to late Pleistocene by recurrent introgressions from its parental species *P. wilsonii* and *P. likiangensis* (Ru et al. 2018).

Table 6 Estimates of nucleotide diversity (π and Θ), Tajima's D, Fay and Wu's H and demographic inferences for 20 *Picea* spp.[a]

Continent/ Species	Number of genes	π	Θ	D	H	Demographic inferences	References
North America							
P. breweriana	10	0.0009	0.0008	0.74	−0.45	Ancient decline; stable small populations afterwards	Chen et al. (2010)
P. breweriana	na (GbS data)[a]	0.0010	0.0007	na	na	–	Haselhorst et al. (2019)
P. engelmannii	~15,000	0.0014	na	na	na	–	Conte et al. (2017)
P. engelmannii x P. glauca	na (GbS data)[a]	0.0097	0.0091	na	na	–	Haselhorst et al. (2019)
P. glauca	10	0.0053	0.0053	−0.36	−0.76	Early Pleistocene decline followed by population growth	Chen et al. (2010)
P. glauca	5	0.0002	0.0010	−0.88	na	Late Pleistocene bottleneck followed by Holocene expansion	Namroud et al. (2010)
P. glauca	105	0.0013	0.0012	−0.40	na	Late Pleistocene bottleneck followed by Holocene expansion	Pavy et al. (2012b)
P. glauca	~15,000	0.0012	na	na	na	–	Conte et al. (2017)
P. glauca	na (GbS data)[a]	0.0090	0.0081	na	na	–	Haselhorst et al. (2019)
P. mariana	10	0.0032	0.0037	−0.69	−0.78	Early Pleistocene decline followed by population growth	Chen et al. (2010)
P. mariana	5	0.0003	0.0008	−0.90	na	Late Pleistocene bottleneck followed by Holocene expansion	Namroud et al. (2010)
P. mariana	na (GbS data)[a]	0.0010	0.0010	na	na	–	Haselhorst et al. (2019)
P. pungens	na (GbS data)[a]	0.0010	0.0007	na	na	–	Haselhorst et al. (2019)

(continued)

Table 6 (continued)

Continent/ Species	Number of genes	π	Θ	D	H	Demographic inferences	References
P. rubens	na (1.3 M SNPs)	0.0050, 0.0054, 0.0054[b]	0.0052, 0.0046, 0.0044[b]	−0.06, 0.19, 0.15[b]	na	Steady Pleistocene decline, stable effective population size for core populations and likely introgression from *P. mariana* in the north	Capblancq et al. (2020)
P. sitchensis	153	0.0012	0.0015	−0.56	−0.36	Historical bottleneck combined with various cycles of expansion-contraction; isolation by distance	Holliday et al. (2010a)
P. sitchensis	na (GbS data)[a]	0.0010	0.0007	na	na	–	Haselhorst et al. (2019)
Europe							
P. abies	22	0.0021	0.0032	−0.92	−0.74	Ancient divergence of two "varieties" (Baltic and Alpine); early Pleistocene bottleneck	Heuertz et al. (2006)
P. abies	10	0.0038	0.0058	−0.90	−0.78	Early Pleistocene decline followed by population growth	Chen et al. (2010)
P. abies	5	0.0004	0.0007	−1.49	na	Late Pleistocene bottleneck followed by Holocene expansion	Namroud et al. (2010)
P. abies	22	0.0039	0.0039	−0.13	na	Ancient divergence from *P. obovata* (Pliocene-Pleistocene) followed by late Pleistocene domain split	Tsuda et al. (2016)
P. abies	~13,000	0.0050	na	−0.85	na	–	Bernhardsson et al. (2019)
P. abies	~7,500	0.0072, 0.0073, 0.0079[c]	na	−0.32, −0.42, 0.34[c]	na	Ancient bottleneck prior to domain splitting, recent bottleneck at the end the Pleistocene	Chen et al. (2019a)
P. obovata	22	0.0033	0.0040	−0.54	na	Ancient divergence from *P. abies* (Pliocene-Pleistocene) followed by recurrent hybridization events	Tsuda et al. (2016)
P. obovata	~7,500	0.0077	na	−0.18	na	Ancient bottleneck (beginning of LGM)	Chen et al. (2019a)

(continued)

Table 6 (continued)

Continent/ Species	Number of genes	π	Θ	D	H	Demographic inferences	References
P. omorika	~7,500	0.0066	na	0.88	na	Very recent bottleneck and sudden contraction of population range	Chen et al. (2019a)
Asia							
P. asperata	13	0.0026	0.0033	0.66	−0.99	Decline since the last interglacial; ancient introgression from *P. crassifolia*	Bi et al. (2016)
P. brachytyla var. brachytyla	~11,000	0.0021	na	na	na	Recent expansion	Ru et al. (2016)
P. brachytyla var. complanata	~10,000	0.0031	na	na	na	Recognized as a *P. linkiagensis* variety recently introgressed by *P. brachyta* var. *brachytyla*	Ru et al. (2016)
P. crassifolia	13	0.0030	0.0039	0.93	−1.02	Decline since the last interglacial	Bi et al. (2016)
P. likiangensis var. likiangensis	13	0.0033	0.0040	−0.31	−0.61	Pliocene divergence and expansion	Li et al. (2013)
P. likiangensis var. rubescens	13	0.0055	0.0044	−0.63	−0.29	Early Pleistocene decline; late Pleistocene expansion	Li et al. (2013)
P. likiangensis var. lizhiensis	13	0.0032	0.0031	−0.38	−0.73	Early Pleistocene decline; late Pleistocene expansion	Li et al. (2013)
P. likiangensis	~14,000	0.0041	na	−0.89	na	Diverged from *P. wilsonii* ~10 Mya	Ru et al. (2018)
P. morrisonicola	15	0.0015	0.0015	0.28	−2.82	A likely derivative from *P. wilsonii*; strong and still ongoing population decline	Bodare et al. (2013), also Li et al. (2010), Zou et al. (2013)
P. neoveitchii	13	0.0026	0.0026	0.77	−0.75	Early Pleistocene expansion; late Pleistocene decline	Bodare et al. (2013), also Wang et al. (2016)
P. purpurea	16	0.0100	na	−0.80	−0.68	Putative hybrid origin from *P. wilsonii* and *P. linkiagensis* (probably in different contact regions); Pleistocene expansion	Li et al. (2010), also Sun et al. (2014)

(continued)

Table 6 (continued)

Continent/ Species	Number of genes	π	Θ	D	H	Demographic inferences	References
P. purpurea	~12,000	0.0039	na	−1.22	na	Hybrid origin from *P. wilsonii* and *P. linkiagensis*	Ru et al. (2018)
P. smithiana	11	0.0017	0.0015	na	na	Pliocene divergence with gene flow from *P. schrenkiana*; stable populations afterwards	Li et al. (2015)
P. schrenkiana	11	0.0012	0.0014	na	na	Pliocene divergence with gene flow from *P. smithiana*; stable populations afterwards	Li et al. (2015)
P. wilsonii	16	0.0085	na	−0.38	0.80	Early Pleistocene expansion followed by late and still ongoing decline; introgression from *P. neoveitchii* in the contact zone	Sun et al. (2014), also Zou et al. (2013), Wang et al. (2016)
P. wilsonii	~12,000	0.0039	na	−1.11	na	Diverged from *P. likiangensis* ~10 Mya	Ru et al. (2018)

[a]*na* not available, *GbS* Genotyping-by-Sequencing
[b]Estimated for Core, Margin, and Edge populations, respectively
[c]Estimated for Alpine, Carpathian, and Fennoscandian domains, respectively. Estimated for coding sites for which all changes were synonymous

Contrary to the popular belief that most demographic events in conifer populations occurred during or after the LGM, most population size changes detected in *Picea* spp. largely predate that period. The general demographic trends inferred so far from analyzing haplotype and nucleotide diversity in spruce species include, among others, modern species' origins followed by ancient bottlenecks and subsequent population expansions (Table 6). For instance, past size population increases have been dated back as far as to the late Pliocene or the early Pleistocene era for *P. glauca*, *P. likiangensis* var. *likiangensis*, and *P. mariana* (Chen et al. 2010; Li et al. 2013), and strong bottlenecks were inferred for *P. asperata*, *P. crassifolia*, *P. mexicana*, *P. meyeri*, or *P. morrisonicola* during or before the last interglacial period, some 130 kya (Bodare et al. 2013; Jaramillo-Correa et al. 2015; Bi et al. 2016; Wang et al. 2016). Largely distributed temperate and boreal species such as *P. abies*, *P. glauca*, and *P. mariana* further experienced more recent expansions (e.g., Namroud et al. 2010; Pavy et al. 2012b), whereas subtropical species secluded into isolated mountain stands either would have kept a rather constant population size since their decline (e.g., *P. breweriana*, *P. smithiana, P. rubens*; Chen et al. 2010; Li et al. 2015; Capblancq et al. 2020 for core red spruce populations), or would still be suffering ongoing demographic collapse (e.g., *P. asperata*, *P. chihuahuana*, *P. crassifolia*, *P. rubens*, *P. neoveitchii*, and *P. omorika*; Jaramillo-Correa et al. 2006; Wehenkel and Saénz-Romero 2012; Bi et al. 2016; Wang et al. 2016; Capblancq et al. 2020 for red spruce margin and edge populations).

Discordance in demographic inferences among studies often exists in the population genomics literature, and the genus *Picea* is not an exception. For example, nucleotide diversity data from regulatory genes in white spruce, black spruce, and Norway spruce (Namroud et al. 2010), as well as additional data from many other white spruce genes (Pavy et al. 2012b), better fitted the hypothesis of a bottleneck during the LGM with subsequent early Holocene expansion, as generally supported by palynological data. On the other hand, nucleotide diversity data from other genes in Norway spruce indicated a far more ancient population bottleneck followed by subsequent population size increase (Heuertz et al. 2006; Chen et al. 2010). However, the most recent and exhaustive study in Norway spruce to date reported a major bottleneck dating back to the very end of the Pleistocene (Chen et al. 2019a), in agreement with earlier studies (Namroud et al. 2010). Bottlenecks inferred for the North American white spruce and black spruce generally agree with the biogeographic history of their major lineages migrating south of the ice front during the LGM, followed by recolonization and expansion of their natural distributions (Namroud et al. 2010; Pavy et al. 2012b). However, contradictory results reported for Norway spruce may be attributed to differences in populations and lineages sampled, to historical introgression with other species (such as *P. obovata*) that may not have always been taken into account, or to variations in the number of genes investigated (see Chen et al. 2010, 2019a).

In plants and trees, knowledge about linkage disequilibrium (LD) is relevant for the design and use of efficient methods in various types of population genomic studies (Neale and Savolainen 2004). Similarly as for pine genes, resequencing of spruce genes has shown that LD decays rapidly, in most cases within gene limits

(Heuertz et al. 2006; Namroud et al. 2010; Pavy et al. 2012b). For instance, LD was scrutinized in 105 white spruce genes by sequencing a panel of haploid megagametophytes from natural populations, which was further compared with LD in Norway spruce and other Pinaceae (Pavy et al. 2012b). Low levels of LD were generally observed within white spruce genes, with half-decay of LD reached at a distance of less than 100 bp, on average, which was smaller than that observed for pine genes. Similar results were also obtained for Norway spruce. This trend has important implications for population genomics studies. If this trend is common throughout the gene space or the genome, a very large number of SNPs should be required to uncover exhaustively phenotypic-genetic associations in natural spruce populations. Thus, from several hundreds of thousands to nearly a million SNPs would be necessary to cover adequately the spruce gene space in exome-wide association studies, assuming an average size of 3–3.5 kb per gene (Hamberger et al. 2009) and ~32,500 transcribed genes in the spruce genome (Rigault et al. 2011). Even more SNPs or other types of variants would need to be screened for association mapping targeting noncoding DNA which represents most of the spruce nuclear genomes (Nystedt et al. 2013; De La Torre et al. 2014a). Hence, *a priori* functional genomics information could be useful in several association study contexts where large numbers of trees must be considered, so to narrow down the lists of target regions, or candidate genes or gene families suspected to be involved in the genetic control of traits of ecological or economical interest, as documented for several key physiological processes (e.g., Lamara et al. 2016).

Conducting neutrality tests on nucleotide diversity is another frequent use of spruce population resequencing data. A full description of these various tests has been provided by Vitti et al. (2013). Briefly, frequency spectrum-based methods such as Tajima's D or Fay and Wu's H are standard procedures to detect molecular signatures of selection in genes or genomic regions of interest, leading to potent markers to be used to monitor adaptive genetic variation in natural or breeding populations and for the conservation of natural genetic resources. Tajima's D test verifies whether the proportion of rare variants in a given sequence meets neutral expectations, while Fay and Wu's H test focuses on the proportion of high frequency derived alleles and determines the ancestral state using an outgroup. These two tests are often reported together, given that H is less sensitive than D to putative confounding effects of demography on selection signals. While near null D and H values are expected under neutrality, negative values of D are generally associated with positive selection, selective sweeps or population expansion, and negative values of H are generally associated with selective sweep or ancient bottleneck. A key way to distinguish between selection versus demographic signatures is that the latter can usually be detected genome-wide.

Negative average D and H values appear to be the rule for most spruce species (Table 6). This trend was expected given that spruce species experienced cycles of range contractions and expansions driven by Pleistocene climatic oscillations, generally associated with population demographic fluctuations. However, evidence for positive and balancing selection, or selective sweeps, was reported in several spruce species for a number of genes (Namroud et al. 2010; Pavy et al. 2012b;

Bernhardsson et al. 2019). In some cases, these signatures of selection were supported by higher than average LD values within the genes of interest (Namroud et al. 2010). The genes with such signatures could thus represent valuable candidates for association studies.

7 Genomic Selection in Spruce Breeding Programs

Genomic selection (GS; Meuwissen et al. 2001), also known as genomic prediction, is a whole-genome regression approach to analyze complex traits in living organisms and predict the worth of individuals making up a breeding population based on their multilocus genomic profiles and their genomic-estimated breeding values (GEBVs). The approach is typically population genomics-based but mostly applied in the context of domesticated or breeding populations. It has the potential to significantly increase the intensity with which favorable alleles can be accumulated in a breeding population, reduce considerably the time needed to complete selection cycles, improve several traits of interest simultaneously including adaptive traits and in turn, increase the gain per time unit, especially for traits of low heritability (Park et al. 2016). The rationale for GS is that a large set of markers evenly distributed across the genome and genotyped within a population of artificially reduced effective size (such as spruce advanced-breeding populations) will capture by virtue of LD most of the QTLs involved in the quantitative traits of interest (Meuwissen et al. 2001; Lebedev et al. 2020).

The potential for use of GS in spruce breeding was recently explored in many ways. Results of 15 studies have so far been published, including six studies on white spruce, four on Norway spruce, three on interior spruce, one on black spruce as well as one on Sitka spruce (summarized in Table 7). These studies considered with success a large number of traits including growth, wood quality, and resistance to weevil attacks and to spruce budworm (Table 7). These proof of concept studies relied on a variety of population types, statistical approaches, and different genotyping platforms. While GS studies in interior spruce relied on 15,000–50,000 markers obtained from genotyping-by-sequencing (GbS), more than 100,000 markers from exome sequencing have been considered for the construction of GS models in Norway spruce (Table 7). For white spruce, black spruce, and Norway spruce, Illumina Infinium iSelect arrays have been used to genotype trees for roughly 4,000–7,000 validated SNPs representing as many distinct gene loci (Table 7). Hence, all spruce GS studies were based on SNP markers, except for the Sitka spruce study, which was carried out using RAD markers (Fuentes-Utrilla et al. 2017).

Several whole-genome-regression approaches of parametric or semi-parametric nature provide estimates of each marker effect (Meuwissen et al. 2001; Gianola et al. 2006; de los Campos et al. 2009). However, the estimation of many thousand marker effects generated by the use of dense SNP panels requires specific algorithms. In forestry, Bayesian approaches have been repeatedly used such as Bayesian least

Table 7 Genomic selection studies in *Picea* spp.

Species	Breeding population[a]	Traits[b]	Population size parents/ families/ progenies	Genotyping methods[c]	Number of SNPs	GS methods[d]	Accuracies [methods][e]	References
P. glauca	Full-sibs	Growth, wood density, MFA	39/59/1,748	Infinium iSelect gene SNP array	6,932	Bayesian RR-BLUP, Bayesian LASSO	0.52–0.79 [TBV-ABLUP]	Beaulieu et al. (2014a)
P. glauca	OP Half-sibs	Growth, wood density, MFA, MOE, cell features	620[f]/214/1,694	Infinium iSelect gene SNP array	6,385	Bayesian RR-BLUP, BayesCπ	0.33–0.44 [TBV-ABLUP]	Beaulieu et al. (2014b)
P. glauca	OP Half-sibs	Growth, wood density	620[f]/214/1,694	Infinium iSelect gene SNP array	6,385	GBLUP	0.65–0.70 [TBV-ABLUP]	Gamal El-Dien et al. (2016)
P. glauca	OP Half-sibs	Growth, wood density	620[f]/214/1,694	Infinium iSelect gene SNP array	6,385	ss-GBLUP	0.47–0.68 [theoretical]	Ratcliffe et al. (2017)
P. glauca	Full-sibs	Growth, acoustic velocity for dynamic wood stiffness, resistance to spruce budworm	212/136/1,516	Infinium iSelect gene SNP array	4,148	GBLUP	0.38–0.67 [ScaledH, TBV-GBLUP, TBV-ABLUP]	Beaulieu et al. (2020)
P. glauca	Half-sibs from polycross mating	Growth, wood density, acoustic velocity for dynamic wood stiffness	54/38/892	Infinium iSelect gene SNP array	4,092	GBLUP	0.44–0.74 [ScaledH, theoretical]	Lenz et al. (2020b)
P. glauca × *P. engelmannii*	OP Half-sibs	Growth, wood density, acoustic velocity for dynamic wood stiffness	Unknown/25/ 1,126	GbS and imputation	16,383	RR-BLUP, GRR, GBLUP	0.34–0.64 [TBV-ABLUP]	Gamal El-Dien et al. (2015)
P. glauca × *P. engelmannii*	OP Half-sibs	Growth at several ages	Unknown/25/ 769	GbS and imputation	34,570 – 50,803	RR-BLUP, GRR, BayesCπ	0.31–0.55 [theoretical]	Ratcliffe et al. (2015)
P. glauca × *P. engelmannii*	OP Half-sibs	Growth, wood density	Unknown/25/ 1,126	GbS and imputation	~30,000	GBLUP	0.62–0.70 [TBV-ABLUP, TBV-GBLUP]	Gamal El-Dien et al. (2018)
P. abies	Full-sibs	Growth, wood density, acoustic velocity for dynamic wood stiffness	55/128/1,370	Exome capture, GbS and imputation	116,765	Bayesian RR-BLUP, Bayesian LASSO, RKHS, GBLUP	0.39–0.81 [TBV-ABLUP]	Chen et al. (2018)

(continued)

Table 7 (continued)

Species	Breeding population[a]	Traits[b]	Population size parents/ families/ progenies	Genotyping methods[c]	Number of SNPs	GS methods[d]	Accuracies [methods][e]	References
P. abies	Full-sibs	Growth, Pilodyn for wood density, acoustic velocity for dynamic wood stiffness	55/128/1,370	Exome capture, GbS and imputation	116,765	GBLUP	0.07–0.46 [PA][g]	Chen et al. (2019b)
P. abies	Full-sibs	Growth, wood density, MFA, acoustic velocity for dynamic wood stiffness, resistance to weevil attacks	35/40/726	Infinium iSelect gene SNP array	3,914	GBLUP, Threshold-GBLUP, Bayesian RR, BayesCπ	0.69–0.91 [ScaledH]	Lenz et al. (2020a)
P. abies	OP Half-sibs	Wood density, MFA, MOE, acoustic velocity, and Pilodyn for dynamic wood stiffness	Unknown/62/ 484	Exome capture, GbS and imputation	130,269	GBLUP, RR-BLUP. BayesB, RKHS	0.35–0.46 [ScaledH]	Zhou et al. (2020)
P. mariana	Full-sibs	Growth, wood density, MFA	27/34/734	Infinium iSelect gene SNP array	4,993	Bayesian RR-BLUP	0.74–0.86 [TBV-ABLUP]	Lenz et al. (2017)
P. sitchensis	Full-sibs	Growth, bud burst	2/1/622	RAD sequencing	8,397	GBLUP	0.40–0.58 [ScaledH]	Fuentes-Utrilla et al. (2017)

[a]*OP* open-pollinated

[b]*MFA* wood microfibril angle, *MOE* wood modulus of elasticity

[c]*GbS* genotype-by-sequencing, *RAD* restriction-site associated DNA

[d]*RR-BLUP* ridge regression best linear unbiased prediction, *LASSO* least absolute shrinkage and selection operator, *GRR* generalized ridge regression, *BayesCπ* Bayesian method with a fraction (1-π) of markers for which the effects are estimated from data, *GBLUP* genomic best linear unbiased prediction

[e]Different measures of accuracy are used in the literature, given that the true breeding values are unknown: "TBV-ABLUP," the product-moment correlation between predicted and the presumably true breeding values obtained from ABLUP on the entire dataset; "TBV-GBLUP," the product-moment correlation between predicted and the presumably true breeding values obtained from GBLUP on the entire dataset; "Theoretical," a theoretical accuracy based on the standard errors of breeding value estimates, see Ratcliffe et al. (2017) for formula; "ScaledH," scaling the predictive ability of phenotypes by the square root of heritability

[f]Effective population size

[g]Predictive Ability (PA) as the correlation between the phenotype and GEBV or else EBV

absolute shrinkage and selection operator (LASSO), Bayesian estimation of variances (BayesCπ), and Bayesian ridge regression (RR-BLUP). Several of these methods assume non-normal distribution of QTL effects and hence, allow marker contribution to overall genetic variance to vary among loci. For example, LASSO assumes a Laplace distribution of marker effects and was tested in datasets of white spruce (Beaulieu et al. 2014a) and Norway spruce (Chen et al. 2018) (see also Table 7). BayesCπ allows modelling a proportion of markers (1–π) to have an effect, while a proportion π of marker effects are shrunk toward zero in order to reflect the expected genetic architecture of traits controlled by few loci of major effect (Ratcliffe et al. 2015 for interior spruce; Lenz et al. 2020a for Norway spruce). Whereas RR-BLUP assumes marker effects to be normally distributed and hence of nonzero effect, which is appropriate for traits controlled by a large number of genes of small effects. This appears to well reflect the known genetic control of wood and growth traits that have been so far assessed in successful proof of concept studies in spruce breeding populations: e.g., white spruce (Beaulieu et al. 2014a), interior spruce (Gamal El-Dien et al. 2015; Ratcliffe et al. 2015), black spruce (Lenz et al. 2017), and Norway spruce (Zhou et al. 2020). Indeed, quite similar GS model accuracies were reported and there seems to be no clear advantage of one or the other statistical approach in the analyses of current spruce datasets with shallow pedigree and relatively low marker densities (Beaulieu et al. 2020; Lenz et al. 2020a).

In recent years, GS modelling based on the genomic realized relationships between population individuals (VanRaden 2008) was more popular because of its ease of implementation and interpretation. Although this method assumes that a priori marker effects are normally distributed with a common variance, it provides results similar to those from other methods with different marker distributions (Lenz et al. 2020b). With this method called GBLUP, GEBVs are estimated using best linear unbiased prediction in a standard mixed model framework (BLUP: Henderson 1975). Based on multilocus genomic profiles, the G-matrix codes for the realized genomic relationships between individuals of a population and thereby replaces the numerator relationship matrix (A-matrix), which represents the registered pedigree information that has traditionally been used in quantitative genetics analysis. However, contrary to the A-matrix, the G-matrix accounts for Mendelian sampling effects resulting from the random segregation of parental alleles (Hill and Weir 2011). Hence, replacing the A-matrix by the G-matrix does not alter standard quantitative genetic analyses, and it confers a major advantage for forward selection because it is possible to predict individual breeding values and not only the family breeding values, as it is the case for the pedigree-based method (ABLUP).

The ease of including realized relationship matrices in mixed modelling has led to further developments to take advantage of all the pedigree information available in breeding populations. Legarra et al. (2009) proposed to combine the pedigree and the genomic information in a blended hybrid relationship matrix (H-matrix). Misztal et al. (2009), in a companion paper, developed an efficient computing strategy to obtain solutions to mixed model equations using the H-matrix and estimate corresponding breeding values, which is known as single-step GBLUP (or rarely

HBLUP). Applying this approach to the white spruce open-pollinated family data of Beaulieu et al. (2014b), Ratcliffe et al. (2017) concluded that single-step GBLUP is a cost-effective approach that allows for increased accuracy of GS models and reduced bias in heritability estimates with already a small proportion of genotyped trees. These results support the implementation of GS in conifer breeding programs with shallow pedigree and large effective populations.

The efficiency of GS is measured by the accuracy of the models to predict the breeding or genetic values based on cross-validation. However, it should be noted that different measures of accuracy are used by GS proof of concept studies in spruces, hampering easy comparison among studies (see Table 7 for details). Therefore, recent studies provided several accuracy estimates to facilitate comparison (Beaulieu et al. 2020; Lenz et al. 2020a).

Model accuracy in published spruce GS studies varies significantly: lowest accuracies are obtained for models based on half-sib families (Beaulieu et al. 2014b; Ratcliffe et al. 2017 using 214 white spruce open-pollinated (OP) families; Zhou et al. 2020 using 62 Norway spruce OP families, and see also Thistlethwaite et al. 2020 for a comparison between interior spruce and *Pseudotsuga menziesii*), which is due to lacking control over the paternal contributions with resulting relatively large effective population size. Accuracy somewhat increases when reduced numbers of OP families or polycross families are considered (Gamal El-Dien et al. 2015 using 25 interior spruce OP families; Lenz et al. 2020a using 38 white spruce polycross families). These results confirmed simulation studies where effective population size was a key factor affecting GS accuracy (Grattapaglia and Resende 2011). In white spruce, Beaulieu et al. (2014a) conducted a first study addressing the application of GS in full-sib families and showed that higher accuracy was obtained when models were based on controlled crosses and thus, smaller effective population size. Thereafter, other studies based on full-sib families also led to high accuracy estimates, such as in black spruce (Lenz et al. 2017) and Norway spruce (Chen et al. 2018; Lenz et al. 2020a). Because LD decays rapidly in spruce natural populations (Pavy et al. 2012b), current GS models should be applied only to predict GEBVs of trees belonging to the same breeding population used for training models, in order to obtain high prediction accuracy and sizeable genetic gains per time unit. Indeed, prediction accuracies were close to zero when trees of the training set and those of the validation set were not genetically related (Beaulieu et al. 2014a, b; Lenz et al. 2017).

It has been shown that higher GS model accuracy is reached asymptomatically not only with decreasing effective population size of the breeding population analyzed, but with increasing trait heritability, larger size of the model training population and larger marker density up to 20 markers per centimorgan (cM) (Grattapaglia and Resende 2011). For a Norway spruce breeding population of full-sib families and using genotyping-by-exome capture and sequencing, Chen et al. (2018) relied on about 6 markers per cM for GS modelling of growth and wood quality traits, which was about twice as many markers per cM than the marker density used for GS modelling from SNP array-based genotyping of white spruce full-sib families (Beaulieu et al. 2014a; Lenz et al. 2020b), black spruce full-sibs

(Lenz et al. 2017), and Norway spruce full-sibs (Lenz et al. 2020b), which were also assessed for growth and wood quality traits in all cases. No notable differences in GS model accuracies could be observed between these two levels of genome coverage among studies. With black spruce full-sib families, Lenz et al. (2017) further showed that reducing the number of well-distributed and high-quality markers to about 1,000 did not significantly hamper the estimation of the "true" marker effects, suggesting that GS model accuracy was mostly dependent on capturing well the relatedness in the target population (Zapata-Valenzuela et al. 2012; Thistlethwaite et al. 2020).

When the repeated use of a SNP array is not planned, GbS could represent an alternative approach to increase genome coverage so to maximize GS model accuracy, especially in the case where large effective population sizes are implicated in GS modelling, such as with half-sib families (e.g., Ratcliffe et al. 2015). Depending on the success of the GbS approach, imputation methods have been frequently used when the rate of missing data was high. The error rate of the data imputation procedures that are usually used in such case can be quite high (for error rates, see Lenz et al. 2020b; Beaulieu et al. 2020) and it is not clear how this might decrease the accuracy of GS models when data imputation is used extensively. It is likely that when LD is decreased in populations of half-sibs with large effective population size as compared to full-sibs, GS accuracy could be more affected by the error rate in data imputation. More studies are needed at this level.

It has also been suggested that strategies focusing on markers having previously shown to harbor large effects may improve GS model accuracy (Lenz et al. 2017; Chen et al. 2018). Given the high level of LD in breeding populations, such ability may not only be related to tracing relatedness but also to tracing recombination hotspots, rather than tracing quantitative trait nucleotides (QTNs) at shorter physical distance from markers. However, such an enrichment strategy should be used with caution, given that estimates of marker effects provided by GS models are not orthogonal and consequently may not represent well the percentage of variance explained as that estimated in conventional single-marker association studies.

Focus traits of GS modelling in spruces have so far mainly been growth and wood quality. However, recent studies also looked at traits related to biotic stress resistance such as white pine weevil attack (Lenz et al. 2020a) or acetophenone aglycones in spruce needles, that are tightly linked to spruce budworm resistance (Beaulieu et al. 2020). Given that large number of candidate trees can be screened with GS, these studies also assessed multi-trait prediction and evaluated the ability to use genetic covariance among traits to predict genomic breeding values when targeted traits cannot be measured on the full population due to operational or financial constraints. Lenz et al. (2020a) showed that strong genetic correlation with a trait with full records can be used to obtain high accuracy on a secondary trait with less complete records, such as data related to pest resistance which are often more scarce or expensive to collect.

Similar to conventional modelling, GS needs to account for genotype-by-environment interaction (GxE), especially when validating models across very different environments (Gamal El-Dien et al. 2015). For example in white spruce, significantly lower across-site prediction accuracy was reported for growth traits than for

wood quality traits (Beaulieu et al. 2014a), indicating higher GxE for growth traits. Gamal El-Dien et al. (2018) highlighted that modelling strategies considering several sites allow for variance partition and including GxE in the model itself. In cases of more uniform environmental variation or traits less sensitive to GxE such as wood quality, GS accuracy remained high even with multi-site modelling (Beaulieu et al. 2014a, 2020; Lenz et al. 2017).

GBLUP and generally all approaches based on the realized genetic relationships G-matrix appear to be more precise than pedigree-based approaches given that pedigree-based estimates of additive genetic variance are generally upwardly biased because of lack of consideration of Mendelian sampling and possible pedigree contaminations. Hence, a trend being well supported in various species (Beaulieu et al., in preparation) indicates that GBLUP and other approaches based on the G-matrix and pedigree reconstruction provide more realistic estimates of genetic parameters such as heritability, thus resulting in less inflated genetic gain estimates than those obtained with the conventional pedigree-based approach (see in Beaulieu et al. 2014a, b; Gamal El-Dien et al. 2016; Lenz et al. 2017, 2020b). This has important implications for the efficiency of conventional spruce breeding programs.

Non-additive genetic effects including dominance and first-order epistatic interactions (including additive-by-additive, dominance-by-dominance, and additive-by-dominance) as well as marker-by-environment interaction (M × E) effects were also recently assessed in Norway spruce full-sib families (Chen et al. 2019b). GS models partitioned additive and non-additive genetic variances more precisely than the pedigree-based models and the predictive ability in GS was substantially increased by including dominance. It was also slightly improved by including M × E effects. Dominance effects were also considered for inclusion in GS modelling in white spruce full-sib and polycross families (Lenz et al. 2020b) and in Norway spruce full-sibs from the Quebec breeding program (Lenz et al. 2020a), but no significant dominance effects could be detected for any of the growth, wood quality, or pest resistance traits analyzed, precluding the inclusion of this effect in GS models. This was likely due to small family size (Lenz et al. 2020b). In a study focusing on GS for a New Brunswick white spruce breeding population of full-sib families (Beaulieu et al. 2020), dominance effects were significant for a few traits and could be included in GS modelling but the accuracy of GBLUP and ABLUP models was not increased, as similarly reported for other tree species (Bouvet et al. 2016; de Almeida Filho et al. 2016). It was argued by de Almeida Filho et al. (2016) that dominance effects have to be large ($d^2 > 0.2$) to improve prediction accuracy.

The research of less than 10 years has documented in many ways the usefulness of GS in spruce breeding programs, which are characterized by especially long breeding cycles. A few general trends are emerging: (1) GbS and SNP arrays are efficient methods to genotype trees, but the later resorts less to data imputation and offers the opportunity to genotype additional trees for the exact same gene marker set, which could be useful for repeated implementation of GS in specific spruce breeding programs or populations; (2) relatedness appears to be the main driver of GS prediction accuracy; consequently, prediction accuracy is generally much higher for full-sibs or polycross progeny than for populations of open-pollinated half-sib

families, which usually harbor much larger effective population sizes. In small advanced-breeding populations characterized by controlled crosses, and as long as both training and testing sets are genetically related, only a few thousand well-distributed Mendelian markers with little missing data are needed to obtain moderate to high prediction accuracy; (3) prediction accuracy can be influenced by GxE and thus, GS modelling should test this source of variation so to optimize breeding strategies; and (4) genetic parameters and genetic gains estimated using marker-based approaches are generally more accurate than those estimated using the conventional pedigree-based approach because of their capability to capture the realized relationships between individuals. Ultimately because large numbers of candidate trees can be assessed by GS without the need for phenotyping, associated costs can be saved, selection can proceed at a very early age from screening young material, multi-trait selection is facilitated and selection intensities can be largely increased, delivering larger gains per generation or time unit than conventional selection.

8 Conservation of Endangered Taxa and Assisted Migration and Gene Flow

Most conservation programs set worldwide are designed for preserving entire habitats or species diversity within habitats. This approach is particularly relevant for taxa confined to a few tractable populations, especially if results from studies of geographical structure are integrated into dynamic conservation strategies. The starting point for such actions is including most of the genetic variation of the focus species into a conservation framework, after either estimating the contribution of each population to the total species genetic pool and/or by identifying putative management units from ecological, geographical, or phylogeographic surveys. This last type of initiatives has the advantage that they provide information about ancient population vicariance and demography, which can be used as a proxy to define conservation units aiming to preserve macro-evolutionary processes and species adaptive potential. Such information is already available for a great variety of threatened spruce species from México (e.g., Ledig et al. 1997, 2000a, b, 2002; Jaramillo-Correa et al. 2006), Asia (e.g., Deshun et al. 2006; Aizawa et al. 2008, 2015; Katsuki et al. 2011), or the Balkans (e.g., Nasri et al. 2008; Aleksić and Geburek 2014). In addition, it is also available for more widely distributed species that are not currently endangered but may have valuable threatened populations or may become vulnerable under future climate shifts, as shown in red spruce (Jaramillo-Correa et al. 2015; Capblancq et al. 2020).

However, phylogeographic surveys have the downside of being conducted with neutral markers at large spatial scales, and at a single point in time (i.e., they seldom cover more than one generation of adult trees). Thus, their results may be unsuitable for evaluating micro-evolutionary processes at the landscape level or for forecasting future evolutionary trends. Local population genetic or genomic surveys testing for

stochastic forces within specific conservation units (or populations) have thus to be performed, such as those that have detected genetic erosion in the younger cohorts of the southern populations of *P. chihuahuana* (Wehenkel and Saénz-Romero 2012) and a decreased number of pollen donors in peripheral stands of *P. sitchensis* (Mimura and Aitken 2007). To the best of our knowledge, no formal micro-scale landscape genetics or genomics studies (i.e., with intensive sampling of trees at distances in the scale of a few meters from each other) have been performed in *Picea* spp., other than those on *P. chihuahuana* (e.g., Wehenkel and Saénz-Romero 2012; Quiñones-Pérez et al. 2014), which were mostly focused on detecting IBD and local-scale genetic structure in relict populations. However, establishing such a landscape genetics or genomics framework for guiding conservation, reforestation and management programs may be useful, as there is growing evidence that natural selection can play a significant role in shaping genetic structure at the micro-site level in forest trees, particularly when driven by variations in soil or other micro-site aspects (e.g., Eckert et al. 2012; Orsini et al. 2013; Forester et al. 2016).

Predicting the adaptive potential of populations or their possible response to specific selective forces such as pest outbreaks, influx of contaminants, or climate change, also relies on detecting the genetic variants or measuring the right quantitative traits involved in these responses. Although a growing number of studies report on genes potentially associated with environmental variation (mainly climate) and their related phenotypes in spruce (see Sects. 4 and 5 above for further details), this information has not been yet integrated into conservation planning. However, constant improvement in phenotypic prediction from multilocus genotypes in various spruce species (see Sect. 7 above), along with reliable environmental zonation derived from quantitative trait variation (see Rodríguez-Quilón et al. 2016 for an example in *Pinus*), promises for integration of population genomics-based predictions into conservation programs in a near future.

The estimation of evolutionary potential and the implementation of population genomics-based predictions seem particularly urgent for subtropical or mountain-isolated spruce taxa, for which other strategies than preserving ongoing evolutionary processes may be necessary. One such action is assisted migration, either in the form of pollen transported from a donor to a target population (assisted gene flow) or by translocating seedlings into new habitats that should become more suitable under future climate change (assisted migration). In the case of the endangered subtropical species *P. chihuahuana*, it has been proposed to establish plantations beyond the species natural ranges (Ledig et al. 2010). For boreal taxa, proposals mostly involve assisted gene flow, in order to increase the frequency of adaptive alleles faster than under natural gene flow. Such strategies have been proposed for white spruce (Andalo et al. 2005; Benomar et al. 2016, 2018; Otis Prud'homme et al. 2018), Sitka spruce (Aitken et al. 2008), and more recently Norway spruce (Milesi et al. 2019). However, caution should be taken as most predictions for assisted migration are mostly based on climate variables, and thus ignore other key factors that may affect plantation/migration success, such as soil properties and biotic interactions such as competing species, pests, or mycorrhizal communities. A more complete evaluation of the realized multi-dimensional niche of each species and the use of

population genomics-based selection strategies to identify pre-adapted individuals (or pollen donors) thus seem necessary to implement such programs (Aitken and Bemmels 2016). Moreover, climate change might bring more extreme and frequent drought conditions (Depardieu et al. 2020, 2021) and generally more unpredictable conditions, such as the recent cooling trend reported for central and eastern North America as well as central and eastern Eurasia, due to the deglaciation of the Arctic Ocean and ensuing shifts of the polar vortex under more meridional latitudes (Kug et al. 2015; Kim et al. 2017).

9 Monitoring of Genetic Diversity and Traceability in Spruce Breeding Programs

Recent advances in genomic information and technology have allowed developing several genotyping applications aimed at monitoring genetic diversity and ensuring traceability in spruce breeding programs. One example aimed at comparing genetic diversity for more than a thousand gene loci between white spruce base undomesticated populations and selected subpopulations for growth improvement under various selection intensities (Namroud et al. 2012). An SNP array designed to target over ~1,000 gene loci was used to genotype the various populations. No alleles were found lost in the selection populations, and only a handful of loci showed a significant shift in allele frequencies between the base population and selected subpopulations. This result was anticipated, given previous studies in white spruce revealing the genetic control of growth traits by many QTLs each with small effect (Pelgas et al. 2011) and thus, weak selection pressures on individual loci. Another example of such applications was recently provided by Galeano et al. (2021), based on genotyping data obtained for ~1,000 white spruce progeny and ~5,000 gene SNPs. They reported that the achievement of 5% genetic gain in tree height through eliminating two-thirds of an open-pollinated seed orchard would result in an eightfold loss in effective population size, but with no significant effect of tree selection on observed heterozygosity and inbreeding index.

A second application in white spruce breeding programs aimed at monitoring the actual level of inbreeding in intra-population crosses by estimating kinships using a genomic relationship matrix obtained from genotyping with a dense gene SNP array. Kinships and inbreeding levels could be evaluated precisely, with significant inbreeding and depressed height growth observed in the progeny (Doerksen et al. 2014).

Another straightforward promising application from using SNP genotyping arrays has been the development of a genetic traceability system for the Québec white spruce advanced-breeding program (Godbout et al. 2017). The integrated system aims to trace the pedigree of genetic materials through all phases of breeding and clonal propagation, including somatic embryogenesis and ensuing production of emblings, so as to ensure the delivery of predicted genetic gains. Therefore, a small

SNP chip (Sequenom iPLEX, ~40 SNPs) was developed by selecting only highly informative markers for a targeted breeding population, allowing for cost-effective periodical assessments of identity and pedigree for thousands of sampled genets and trees (Godbout et al. 2017). The first large screening of cryopreserved cell lines and other clonal material established in the field indicated that 8% of samples had pedigree errors. Pollen contamination at the stage of crossing the selected parents appeared as the main cause for these discrepancies, followed by errors incurred during clonal propagation. In another recent study that relied on using pedigree reconstruction from genotyping for a few thousand SNPs, Galeano et al. (2021) showed that, on average, 30% of seedlots were contaminated by exogenous pollen in a white spruce open-pollinated seed orchard. The main source of pollen contamination was located 1 km away from the seed orchard upstream of prevailing winds. Similar findings using microsatellites were aso reported in *P. abies* (Sønstebø et al. 2018).

10 Future Perspectives

10.1 Toward Integrating Population Genomics with Other "Omics"

In order to bypass difficulties associated with whole-genome sequencing, considerable efforts have been deployed to explore the spruce gene space, which likely contains much of the functionally relevant information. The sequencing of several spruce transcriptomes, followed by exomes, has been the source of major spruce annotated genomic resources including sequence and SNP data, as well as gene expression data. These resources can be exploited in various fields of research and offer many integrative and comparative genomics opportunities.

For instance, single-copy gene sequences were recently used in a phylogenomics context to resolve phylogenetic relationships among gymnosperm lineages (Li et al. 2017). Such approaches will undoubtedly be expanded to more taxa in the near future and help comprehend the evolution of the Pinaceae family and the genus *Picea* in particular. Linking gene and species phylogenies will also be a major area of research, as resequencing efforts are expected to increase for a larger number of spruce taxa. As shown recently in white spruce by Depardieu et al. (2021) and Laoué et al. (2021), the integration of population genomics with expression studies should also intensify given the large high-quality expression resources that have been developed. While hastening the validation of outcomes from association studies (e.g., Chen et al. 2012; Prunier et al. 2015), this integration of approaches should help identify genes and gene families with high rates of substitution and tissue-specific expression (e.g., Beaulieu et al. 2011; Pavy et al. 2013b; De La Torre et al. 2015), thus providing useful lists of candidate genes to be tested in association studies. In addition, such studies may help identify polymorphisms in genes' *cis*-

regulatory regions, which should contribute to bridge the gap between population genomics and transcriptomics and pave the way for subsequent fundamental studies and field applications.

Large catalogs of high-confidence and annotated gene SNPs representing much of the spruce transcriptome and exome have been released (Pavy et al. 2013a, 2016; Azaiez et al. 2018), and they should continue to provide opportunities to further densify white spruce and Norway spruce reference genetic maps (Pavy et al. 2017; Bernhardsson et al. 2019, respectively), and expand other spruce genetic maps to test for micro-collinearity and identify the frequency of segmental rearrangements. This will also result in more precise QTL dissection, provided that large populations are genotyped and phenotyped. Comparing the genomic architecture of quantitative traits between species should then become possible, assisting in delineating valuable sets of candidate genes for association mapping and ecological genomics studies.

One of the main future challenges lies in the validation of the many results reported in association mapping studies through independent means. One avenue involves determining the functional annotation of candidate genes, as well as surveying their expression in relation to specific pathways and environmental conditions. Much groundwork has already been accomplished in this regard by analyzing genomic, microarray, and RNA-Seq data in several spruce species (e.g., Raherison et al. 2012, 2015; Yeaman et al. 2014; Warren et al. 2015; Jokipii-Lukkari et al. 2017; Trujillo-Moya et al. 2018; Depardieu et al. 2021; Laoué et al. 2021; Whitehill et al. 2021; Zhu et al. 2021). However, many spruce or conifer-specific transcripts and genes remain unannotated to date, and extensive efforts will be required to further improve the annotation of spruce transcriptomes, which will facilitate population genomics studies aiming to decipher genetic adaptation and fitness, what is sometimes called functional population genomics. Validation by assessing gene expression may also benefit from the advent of breakthrough technologies such as CRISPR/cas9 genome editing (Tsai and Xue 2015). In any case, validation of candidate genes or loci identified by population genomics approaches is moving toward integrating several lines of evidence to corroborate inferences. This may include crosstalks among data types (e.g., sequence/polymorphism data, expression data, phenotypic data, functional evidence, etc.), crosstalks among disciplines and analyses (phylogenomics, phylogeography, association studies, linkage mapping, etc.), or a combination of both.

Together with accumulating evidence from GWAS (e.g., Hornoy et al. 2015; Yeaman et al. 2016; Milesi et al. 2019; Depardieu et al. 2021), a more precise picture of adaptation at the population genomics level is emerging for long-lived spruces, which supports the conventional view of complex traits being controlled by multiple genes carrying SNPs under positive selection but with small effects. The increasing large body of literature emerging from spruce association mapping studies also indicates that genes carrying adaptive polymorphisms do not appear to be randomly distributed across the transcriptome, with specific functions, gene families, and metabolic pathways more represented than others. Recent studies also indicate that adaptation to temporarily and spatially heterogeneous environments is likely to be complemented by more complex structural variation such as CNVs (Prunier et al.

2017a, b) and by the expansion of key gene families which are thought to provide more metabolic flexibility in response to biotic and abiotic stress (e.g., Warren et al. 2015; Stival Sena et al. 2018).

There is also recent evidence from large-scale association studies that the genomic architecture of complex traits is not simple, with instances of spatial clustering of significant SNPs for key conifer metabolic traits (De La Torre et al. 2019). Whether this pattern is related to functional gene clusters (Pavy et al. 2017) or to an overrepresentation of significant SNPs in regulatory DNA regions of the open chromatin (Grattapaglia et al. 2018) remains to be clarified. However, this example illustrates well that the number of integrative genomics studies is expected to increase in the future, along with the ever-growing volume of quality conifer genomic data produced. In turn, these resources will improve the understanding of conifer and spruce genome biology, evolution, and structure in various population and evolutionary genomics contexts.

10.2 Toward the Population Genomics Information Age in Spruce Forestry

A major expected trend in the future relates to field applications of spruce genomics research and information. In addition to contributions to fundamental knowledge, several genomics tools of demonstrated efficiency were developed recently in order to inform policy makers and assist conservationists and forestry practitioners (see Sects. 7, 8 and 9 above for further details). Along with studies assessing the distribution of neutral genetic diversity, association mapping studies linking genomic information to phenotypes and environmental conditions represent the starting point for integrating genomic information into genetic conservation programs. The efficiency of in situ conservation strategies could thus be improved by delineating optimal conservation units that will maximize conservation of both neutral and adaptive genetic diversity (Funk et al. 2012). Alternatively, these studies can guide sampling strategies for ex situ conservation efforts by identifying hotspots of genetic diversity and genetically depauperate species in need of specific conservation efforts. In line with this idea, assisted migration stands at the border between forest conservation and forest management strategies. Because of ongoing climate change, it can be seen either as a conservation strategy for endangered species facing extinction if no action is taken (Ledig et al. 2010) or a management strategy that can help improve forest resilience and productivity for species exploited commercially, but that are gradually becoming maladapted to their current local environments because of climate change (Andalo et al. 2005; Aitken et al. 2008; Milesi et al. 2019). In either case, population genomics provides valuable insights regarding the genetic background of trees, and can therefore help forest practitioners selecting trees that will be best adapted to their future environment.

Applications of genomics in the management of domesticated genetic resources should also take off, given the obvious benefits of better tracing pedigree in breeding and production populations to deliver the predicted genetic gains (e.g., Godbout et al. 2017; Galeano et al. 2021; see also Sect. 9 above for details). The large benefits shown from the implementation of early genomic prediction in spruce advanced-breeding programs should also fuel this expansion (Park et al. 2016). GS is rapidly becoming a standard tool to screen candidate trees at a very early stage and in larger numbers, in order to hasten selection and breeding, increase selection intensity, or facilitate multi-trait selection while preserving genetic diversity in spruce breeding programs.

This trend is likely to be exacerbated given the multiple traits that are becoming the subject of scrutiny by spruce breeders under the contexts of climate change and rapidly evolving forest product markets (see Spruce-Up project, http://spruce-up.ca), as currently seen in crop breeding (e.g., Das Graças Dias et al. 2018). Many of these traits related to adaptation to climate, pest resistance, or wood quality are expensive or difficult to assess on large cohorts of trees, and/or must await tree maturity before selections are being made, thus increasing the need to develop performant genomic prediction tools. This rapidly evolving context also underlines the necessity to pursue actively common-garden studies, which are essential to develop GS and to assist in identifying relevant genetic marker information related to adaptive traits and fitness. Implicating breeders and other users of genomic information at the early stage of the development of genomic resources for a new species should also favor the uptake of genomic information and the sustainable development of genomic research, including the establishment and maintenance of valuable multi-purpose field tests.

11 Conclusions

Conducting population genomics studies in non-model or undomesticated species is often a challenging task due to the lack of genomic resources, particularly in organisms with very large and highly heterozygous genomes such as conifers and spruces in particular. However, this caveat is progressively being overcome in spruce, given that several transcriptomes/exomes and cytoplasmic genomes have been resequenced, and that whole nuclear genome sequences, albeit still fragmented, have been released for a few species and are being actively improved and made more contiguous, together with the development of affordable GbS approaches. These advances have allowed the development of extensive genomic resources, such as large catalogs of annotated genes and SNPs, hybridization microarrays that have resulted in gene expression databases, high-resolution gene linkage maps, and high-throughput genotyping chips for several species. Owing to the amount of genomic data already obtained, as well as the economical, ecological, and societal importance of spruces worldwide, the genus *Picea* has become a reference among conifers for population genomics research.

Field resources gathered in common-garden experiments are essential for phenotypic assessment of trees in homogenous environments, efficient large-scale sampling across species range, and ex situ conservation. The maintenance and expansion of these trials is becoming even so more critical for a number of spruce species, as climate change and deforestation are taking place. Thanks to these trials, an array of phylogeographic studies could characterize the range-wide distribution of neutral genetic variation for many spruce taxa. Spruce geneticists more recently took advantage of this background information to conduct association mapping and ecological genomics studies, linking phenotypic and environmental variation to genetic variation at the molecular level. Insights derived from these population genomic studies are now available to forest practitioners to improve the efficiency of genetic resources conservation and tree breeding efforts.

With the increasing power of sequencing technologies and bioinformatics algorithms, spruce population genomics should benefit from the integrative analysis of large data sets of various nature, such as exome and genome sequences, genetic maps, expression data, as well as metabolic, physiological, and dendrometric data. This should open new perspectives to improve our understanding of evolutionary processes driving the distribution of selectively neutral and fitness-related genetic variation among spruce species and their populations.

References

Abbott RJ, Brochmann C. History and evolution of the arctic flora: in the footsteps of Eric Hultén. Mol Ecol. 2003;12:299–313.

Acheré V, Faivre-Rampant P, Jeandroz S, Besnard G, Markussen T, Aragones A, et al. A saturated consensus linkage map of *Picea abies* including AFLP, SSR, STS, 5S rDNA and morphological markers. Theor Appl Genet. 2004;108:1602–13.

Aitken SN, Bemmels JB. Time to get moving: assisted gene flow of forest trees. Evol Appl. 2016;9:271–90.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl. 2008;1:95–111.

Aizawa M, Yoshimaru H, Saito H, Katsuki T, Kawahara T, Kitamura K, et al. Phylogeography of a northeast Asian spruce, *Picea jezoensis*, inferred from genetic variation observed in organelle DNA markers. Mol Ecol. 2007;16:3393–405.

Aizawa M, Yoshimaru H, Katsuki T, Kaji M. Imprint of post-glacial history in a narrowly distributed endemic spruce, *Picea alcoquiana*, in central Japan observed in nuclear microsatellites and organelle DNA markers. J Biogeogr. 2008;35:1295–307.

Aizawa M, Yoshimaru H, Takahashi M, Kawahara T, Sugita H, Saito H, et al. Genetic structure of Sakhalin spruce (*Picea glehnii*) in northern Japan and adjacent regions revealed by nuclear microsatellites and mitochondrial gene sequences. J Plant Res. 2015;128:91–102.

Aizawa M, Yoshimaru H, Ogawa H, Goto S, Kaji M. Natural hybridization of Yeko and Sakhalin spruce in central Hokkaido, revealed by DNA markers with contrasting modes of inheritance. Plant Species Biol. 2016;31:188–95.

Aleksić JM, Geburek T. Quaternary population dynamics of an endemic conifer, *Picea omorika*, and their conservation implications. Conserv Genet. 2014;15:87–107.

Alsos IG, Eidesen PB, Ehrich D, Skrede I, Westergaard K, Jacobsen GH, et al. Frequent long-distance plant colonization in the changing Arctic. Science. 2007;316:1606–9.

Andalo C, Beaulieu J, Bousquet J. The impact of climate change on growth of local white spruce populations in Québec, Canada. For Ecol Manage. 2005;205:169–82.

Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN. Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. Proc Natl Acad Sci U S A. 2006;103:12447–50.

Arnold ML. Divergence with genetic exchange. Oxford: Oxford University Press; 2016.

Azaiez A, Pavy N, Gérardi S, Laroche J, Boyle B, Gagnon F, et al. A catalog of annotated high-confidence SNPs from exome capture and sequencing reveals highly polymorphic genes in Norway spruce (*Picea abies*). BMC Genomics. 2018;19:942.

Baison J, Vidalis A, Zhou L, Chen Z-Q, Li Z, Sillanpää MJ, et al. Genome-wide association study identified novel candidate loci affecting wood formation in Norway spruce. Plant J. 2019;100:83–100.

Barton NH, Hewitt GM. Analysis of hybrid zones. Annu Rev Ecol Syst. 1985;16:113–48.

Bashalkhanov S, Eckert AJ, Rajora OP. Genetic signatures of natural selection in response to air pollution in red spruce (*Picea rubens*, Pinaceae). Mol Ecol. 2013;22:5877–89.

Beaulieu J, Doerksen T, Boyle B, Clément S, Deslauriers M, Beauseigle S, et al. Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. Genetics. 2011;188:197–214.

Beaulieu J, Doerksen TK, MacKay J, Rainville A, Bousquet J. Genomic selection accuracies within and between environments and small breeding groups in white spruce. BMC Genomics. 2014a;15:1048.

Beaulieu J, Doerksen T, Clément S, MacKay J, Bousquet J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. Heredity. 2014b;113:343–52.

Beaulieu J, Nadeau S, Ding C, Celedon JM, Azaiez A, Ritland C, et al. Genomic selection for resistance to spruce budworm in white spruce and relationships with growth and wood quality traits. Evol Appl. 2020;13:2704–22.

Beaumont MA, Balding DJ. Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol. 2004;13:969–80.

Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. Proc R Soc B. 1996;263:1619–26.

Bedon F, Bomal C, Caron S, Levasseur C, Boyle B, Mansfield SD, et al. Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses. J Exp Bot. 2010;61:3847–64.

Bennett KD, Tzedakis PC, Willis KJ. Quaternary refugia of north European trees. J Biogeogr. 1991;18:103–15.

Benomar L, Lamhamedi MS, Rainville A, Beaulieu J, Bousquet J, Margolis HA. Genetic adaptation vs. ecophysiological plasticity of photosynthetic-related traits in young *Picea glauca* trees along a regional climatic gradient. Front Plant Sci. 2016;7:48.

Benomar L, Lamhamedi MS, Pépin S, Rainville A, Lambert M-C, Margolis HA, et al. Thermal acclimation of photosynthesis and respiration of southern and northern white spruce seed sources along a regional climatic gradient indicates limited potential to cope with temperature warming. Ann Bot. 2018;121:443–57.

Bernhardsson C, Vidalis A, Wang X. An ultra-dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (*Picea abies*). G3 Genes Genom Genet. 2019;9:1623–32.

Bernhardsson C, Zan Y, Chen Z, Ingvarsson PK, Wu HX. Development of a highly efficient 50K single nucleotide polymorphism genotyping array for the large and complex genome of Norway spruce (*Picea abies* L. Karst) by whole genome resequencing and its transferability to other spruce species. Mol Ecol Resour. 2021;21:880–96.

Bi H, Yue W, Wang X, Zou J, Li L, Liu J, et al. Late Pleistocene climate change promoted divergence between *Picea asperata* and *P. crassifolia* on the Qinghai–Tibet Plateau through recent bottlenecks. Ecol Evol. 2016;6:4435–44.

Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, et al. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics. 2013;29:1492–7.

Bodare S, Stocks M, Yang J-C, Lascoux M. Origin and demographic history of the endemic Taiwan spruce (*Picea morrisonicola*). Ecol Evol. 2013;3:3320–33.

Bouillé M, Bousquet J. Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic diversity in trees. Am J Bot. 2005;92:63–73.

Bouillé M, Senneville S, Bousquet J. Discordant mtDNA and cpDNA phylogenies indicate geographic speciation and reticulation as driving factors for the diversification of the genus *Picea*. Tree Genet Genomes. 2011;7:469–84.

Bouvet J-M, Makouanzi G, Cros D, Vigneron P. Modelling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. Heredity. 2016;116:146–57.

Ćalić I, Bussotti F, Martínez-García PJ, Neale DB. Recent landscape genomics studies in forest trees – what can we believe? Tree Genet Genomes. 2016;12:3.

Capador-Barreto H, Bernhardsson C, Milesi P, Vos I, Lundén K, Wu HX, et al. Killing two enemies with one stone? Genomics of resistance to two sympatric pathogens in Norway spruce. Mol Ecol. In press, on-line. 2021; https://doi.org/10.1111/mec.16058.

Capblancq T, Butnor JR, Deyoung S, Thibault E, Munson H, Nelson DM, et al. Whole-exome sequencing reveals a long-term decline in effective population size of red spruce (*Picea rubens*). Evol Appl. 2020;13:2190–205.

Chen J, Källman T, Gyllestrand N, Lascoux M. New insights on the speciation history and nucleotide diversity of three boreal spruce species and a tertiary relict. Heredity. 2010;104:3–14.

Chen J, Källman T, Ma X, Gyllenstrand N, Zaina G, Morgante M, et al. Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). Genetics. 2012;191:865–81.

Chen C, Mitchell SE, Elshire RJ, Buckler ES, El-Kassaby YA. Mining conifers' megagenome using rapid and efficient multiplexed high-throughput genotype-by-sequencing (GBS) SNP discovery platform. Tree Genet Genomes. 2013;9:1537–44.

Chen Z-Q, Baison J, Pan J, Karlsson B, Andersson B, Westin J, et al. Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. BMC Genomics. 2018;19:946. (16 p.)

Chen J, Li L, Milesi P, Jansson G, Berlin M, Karlsson B, et al. Genomic data provide new insights on the demographic history and the extent of recent material transfers in Norway spruce. Evol Appl. 2019a;12:1539–51.

Chen Z-Q, Baison J, Pan J, Westin J, García Gil MR, Wu HX. Increased prediction ability in Norway spruce trials using a marker x environment interaction and non-additive genomic selection model. J Hered. 2019b;110:830–43.

Choat B, Jansen S, Brodribb TJ, Cochard H, Delzon S, Bhaskar R, et al. Global convergence in the vulnerability of forests to drought. Nature. 2012;491:752–5.

Comtois P. Pollen dispersal and long distance transport: the case of thermophilic pollen in subarctic Canada. Aerobiologia. 1997;13:37–42.

Conte GL, Hodgins KA, Yeaman S, Degner JC, Aitken SN, Rieseberg LH, et al. Bioinformatically predicted deleterious mutations reveal complementation in the interior spruce hybrid complex. BMC Genomics. 2017;18:970.

Coombe L, Warren RJ, Jackman SD, Yang C, Vandervalk BP, Moore RA, et al. Assembly of the complete Sitka spruce chloroplast genome using 10X Genomics' GemCode sequencing data. PLoS One. 2016;11:e0163059.

Coop G, Witonski D, Di Rienzo A, Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. Genetics. 2010;185:1411–23.

Cronn R, Liston A, Parks M, Gernandt DS, Shen R, Mockler T. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res. 2008;36:e122.

Das Graças Dias KO, Gezan SA, Guimarães CT, Nazarian A, da Costa e Silva L, Parentoni SN, et al. Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. Heredity. 2018;121:24–37.

Davis MB, Shaw RG. Range shifts and adaptive responses to quaternary climate change. Science. 2001;292:673–9.

de Almeida Filho JE, Guinäraes J, e Silva FF, de Resende MDV, Muñoz P, Kirst M, et al. The contribution of dominance to phenotype prediction in a pine breeding and simulated populations. Heredity. 2016;117:33–41.

De La Torre AR, Birol I, Bousquet J, Ingvarsson PK, Jansson S, Jones SJM, et al. Insights into conifer giga-genomes. Plant Physiol. 2014a;166:1724–32.

De La Torre AR, Roberts DR, Aitken SN. Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. Mol Ecol. 2014b;23:2046–59.

De La Torre AR, Wang T, Jacquish B, Aitken SN. Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: implications for forest management under climate change. New Phytol. 2014c;201:687–99.

De La Torre AR, Lin Y-C, Van de Peer Y, Ingvarsson PK. Genome-wide analysis reveals diverged patterns of codon bias, gene expression and rates of sequence evolution in *Picea* gene families. Genome Biol Evol. 2015;7:1002–15.

De La Torre AR, Puiu D, Crepeau MW, Stevens K, Salzberg SL, Langley CH, et al. Genomic architecture of complex traits in loblolly pine. New Phytol. 2019;221:1789–901.

de Lafontaine G, Bousquet J. Asymmetry matters: a genomic assessment of directional biases in gene flow between hybridizing spruces. Ecol Evol. 2017;7:3883–93.

de Lafontaine G, Turgeon J, Payette S. Phylogeography of white spruce (*Picea glauca*) in eastern North America reveals contrasting ecological trajectories. J Biogeogr. 2010;37:741–51.

de Lafontaine G, Prunier J, Gérardi S, Bousquet J. Tracking the progression of speciation: variable patterns of introgression across the genome provide insights on the species delimitation between progenitor–derivative spruces (*Picea mariana* × *P. rubens*). Mol Ecol. 2015;24:5229–47.

de Lafontaine G, Napier JD, Petit RJ, Hu FS. Invoking adaptation to decipher the genetic legacy of past climate change. Ecology. 2018;99:1530–46.

de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics. 2009;182:375–85.

Depardieu C, Girardin MP, Nadeau S, Lenz P, Bousquet J, Isabel N. Adaptive genetic variation to drought in a widely distributed conifer suggests a potential for increasing forest resilience in a drying climate. New Phytol. 2020;227:427–39.

Depardieu C, Gérardi S, Nadeau S, Parent GJ, Mackay J, Lenz P, et al. Connecting tree-ring phenotypes, genetic associations, and transcriptomics to decipher the genomic architecture of drought adaptation in a widespread conifer. Mol Ecol. 2021;16:3898–917.

Deshun Z, Yongshik K, Maunder M, Xiufen L. The conservation status and conservation strategy of *Picea neoveitchii*. Chinese J Popul Resour Environ. 2006;4:58–64.

Di Pierro EA, Mosca E, Rocchini D, Binelli G, Neale DB, La Porta N. Climate-related adaptive genetic variation and population structure in natural stands of Norway spruce in the South-Eastern Alps. Tree Genet Genomes. 2016;12:16.

Di Pierro EA, Mosca E, González-Martínez SC, Binelli G, Neale DB, La Porta N. Adaptive variation in natural Alpine populations of Norway spruce (*Picea abies* [L.] Karst) at regional scale: landscape features and altitudinal gradient effects. Forest Ecol Manag. 2017;405:350–9.

Doerksen TK, Bousquet J, Beaulieu J. Inbreeding depression in intra-provenance crosses driven by founder relatedness in white spruce. Tree Genet Genomes. 2014;10:203–12.

Du FK, Petit RJ, Liu JQ. More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other conifers. Mol Ecol. 2009;18:1396–407.

Du FK, Peng X-L, Liu J-Q, Lascoux M, Hu FS, Petit RJ. Direction and extent of organelle DNA introgression between two spruce species in the Qinghai-Tibetan Plateau. New Phytol. 2011;192:1024–33.

Eckert AJ, Shahi H, Datwyler SL, Neale DB. Spatially variable natural selection and the divergence between parapatric subspecies of lodgepole pine (*Pinus contorta*, Pinaceae). Am J Bot. 2012;99:1323–34.

Eklöf H, Bernhardsson C, Ingvarsson PK. Comparing the effectiveness of exome capture probes, genotyping by sequencing and whole-genome re-sequencing for assessing genetic diversity in natural and managed stands of *Picea abies*. Forests. 2020;11:1185.

Elfstrand M, Zhou L, Baison J, Olson Å, Lundén K, Karlsson B, et al. Genotypic variation in Norway spruce correlates to fungal communities in vegetative buds. Mol Ecol. 2020;29:199–213.

Farjon A. Pinaceae. Drawings and descriptions of the genera *Abies*, *Cedrus*, *Pseudolarix*, *Keteleeria*, *Nothotsuga*, *Tsuga*, *Cathaya, Pseudotsuga*, *Larix* and *Picea*. Königstein: Koeltz Scientific Books; 1990.

Farjon A. A natural history of conifers. Portland: Timber Press; 2008.

Farjon A. A handbook of the world's conifers. Leiden: Brill Academic Publishers; 2010.

Farjon A, Filer D. An atlas of the world's conifers. Leiden: Brill Academic Publishers; 2013.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Mol Ecol. 2016;25:104–20.

Franklin EC. Survey of mutant forms and inbreeding depression in species of the family Pinaceae. USDA Forest Service Research Paper SE-61, Asheville; 1970.

Fuentes-Utrilla P, Goswami C, Cottrell JE, Pong-Wong R, Law A, A'Hara SW, et al. QTL analysis and genomic selection using RADseq derived markers in Sitka spruce: the potential utility of within family data. Tree Genet Genomes. 2017;13:33.

Funk WC, Mckay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27:489–96.

Galeano E, Bousquet J, Thomas BR. SNP-based analysis reveals unexpected features of genetic diversity, parental contributions and pollen contamination in a white spruce breeding program. Sci Rep. 2021;11:4990.

Gamache I, Jaramillo-Correa JP, Payette S, Bousquet J. Diverging patterns of mitochondrial and nuclear DNA diversity in subarctic black spruce: imprint of a founder effect associated with postglacial colonization. Mol Ecol. 2003;12:891–901.

Gamal El-Dien O, Ratcliffe B, Klápště J, Chen C, Porth I, El-Kassaby YA. Prediction accuracies for growth and wood attributes of interior spruce using genotyping-by-sequencing. BMC Genomics. 2015;16:370.

Gamal El-Dien O, Ratcliffe B, Klápště J, Porth I, Chen C, El-Kassaby YA. Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from non-additive genetic effects. G3 Genes Genom Genet. 2016;6:743–53.

Gamal El-Dien O, Ratcliffe B, Klápště J, Porth I, Chen C, El-Kassaby YA. Multienvironment genomic variance decomposition analysis of open-pollinated interior spruce (*Picea glauca x engelmannii*). Mol Breed. 2018;38:26.

Ganthaler A, Stöggl W, Mayr S, Kranner I, Schüler S, Wischnitzki E, et al. Association genetics of phenolic needle compounds in Norway spruce with variable susceptibility to needle bladder rust. Plant Mol Biol. 2017;94:229–51.

Gérardi S, Jaramillo-Correa JP, Beaulieu J, Bousquet J. From glacial refugia to modern populations: new assemblages of organelle genomes generated by differential cytoplasmic gene flow in transcontinental black spruce. Mol Ecol. 2010;19:5265–80.

Gernandt DS, Magallón S, López GG, Flores OZ, Willyard A, Liston A. Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae phylogeny. Int J Plant Sci. 2008;169:1086–99.

Gernandt DS, Willyard A, Syring JV, Liston A. The conifers (Pinophyta). In: Plomion C, Bousquet J, Kole C, editors. Genetics, genomics and breeding of conifers. New York: CRC Press; 2011. p. 1–39.

Gianola D, Fernando RL, Stella A. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics. 2006;173:1761–76.

Godbout J, Tremblay L, Levasseur C, Lavigne P, Rainville A, Mackay J, et al. Development of a traceability system based on a SNP array for large-scale production of high-value white spruce (*Picea glauca*). Front Plant Sci. 2017;8:1264.

Gordon AG. The taxonomy and genetics of *Picea rubens* and its relationship to *Picea mariana*. Can J Bot. 1976;54:781–813.

Gosselin I, Zhou Y, Bousquet J, Isabel N. Megagametophyte-derived linkage maps of white spruce (*Picea glauca*) based on RAPD, SCAR and ESTP markers. Theor Appl Genet. 2002;104:987–97.

Grattapaglia D, Resende MDV. Genomic selection in forest tree breeding. Tree Genet Genomes. 2011;7:241–55.

Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, Tan B, et al. Quantitative genetics and genomics converge to accelerate forest tree breeding. Front Plant Sci. 2018;9:1693.

Guillet-Claude C, Isabel N, Pelgas B, Bousquet J. The evolutionary implications of *knox-I* gene duplications in conifers: correlated evidence from phylogeny, gene mapping, and analysis of functional divergence. Mol Biol Evol. 2004;21:2232–45.

Hall D, Hallingback HR, Wu HX. Estimation of number and size of QTL effects in forest tree traits. Tree Genet Genomes. 2016;12:110.

Hamberger B, Hall D, Yuen M, Oddy C, Hamberger B, Keeling CI, et al. Targeted isolation, sequence assembly and characterization of two white spruce (*Picea glauca*) BAC clones for terpenoid synthase and cytochrome P450 genes involved in conifer defence reveal insights into a conifer genome. BMC Plant Biol. 2009;9:106.

Hamilton JA, Aitken SN. Genetic and morphological structure of a spruce hybrid (*Picea sitchensis* × *P. glauca*) zone along a climatic gradient. Am J Bot. 2013;100:1651–62.

Hamilton JA, Lexer C, Aitken SN. Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis* × *P. glauca*) hybrid zone. New Phytol. 2013a;197:927–38.

Hamilton JA, Lexer C, Aitken SN. Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* × *P. glauca*). Mol Ecol. 2013b;22:827–41.

Hart JA. A cladistic analysis of conifers: preliminary results. J Arnold Arboretum. 1987;68:269–307.

Haselhorst MSH, Parchman TL, Buerkle CA. Genetic evidence for species cohesion, substructure and hybrids in spruce. Mol Ecol. 2019;28:2029–45.

Heer K, Ullrich KK, Liepelt S, Rensing SA, Zhou J, Ziegenhagen B, et al. Detection of SNPs based on transcriptome sequencing in Norway spruce (*Picea abies* (L.) Karst). Conserv Genet Resour. 2016;8:105–7.

Henderson CR. Best linear unbiased estimation and prediction under a selection model. Biometrics. 1975;31:423–47.

Heuertz M, De Paoli E, Källman T, Larsson H, Jurman I, Morgante M, et al. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. Genetics. 2006;174:2095–105.

Hewitt GM. The genetic legacy of the quaternary ice ages. Nature. 2000;405:907–13.

Hewitt GM. Quaternary phylogeography: the roots of hybrid zones. Genetica. 2011;139:617–38.

Hill WG, Weir BS. Variation in actual relationship as a consequence of Mendelian sampling and linkage. Genet Res. 2011;93:47–64.

Holliday JA, Yuen M, Ritland K, Aitken SN. Postglacial history of a widespread conifer produces inverse clines in selective neutrality tests. Mol Ecol. 2010a;19:3857–64.

Holliday JA, Ritland K, Aitken SN. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). New Phytol. 2010b;188:501–14.

Holliday JA, Wang T, Aitken S. Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (*Picea sitchensis*) using random forest. G3 Genes Genom Genet. 2012;2:1085–93.

Holliday JA, Aitken SA, Cooke JEK, Fady B, González-Martínez SC, Heuertz M, et al. Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. Mol Ecol. 2017;26:706–17.

Hornoy B, Pavy N, Gérardi S, Beaulieu J, Bousquet J. Genetic adaptation to climate in white spruce involves small to moderate allele frequency shifts in functionally diverse genes. Genome Biol Evol. 2015;7:3269–85.

Housset JM, Nadeau S, Isabel N, Depardieu C, Duchesne I, Lenz P, et al. Tree rings provide a new class of phenotypes for genetic associations that foster insights into adaptation of conifers to climate change. New Phytol. 2018;218:630–45.

Huntley B, Webb T III. Migration: species' response to climatic variations caused by changes in the earth's orbit. J Biogeogr. 1989;16:5–19.

Isabel N, Beaulieu J, Bousquet J. Complete congruence between gene diversity estimates derived from genotypic data at enzyme and RAPD loci in black spruce. Proc Natl Acad Sci U S A. 1995;92:6369–73.

Jackman SD, Warren RL, Gibb EA, Vandervalk BP, Mohamadi H, Chu J, et al. Organellar genomes of white spruce (*Picea glauca*): assembly and annotation. Genome Biol Evol. 2016;8:29–41.

Jackman SD, Coombe L, Warren RL, Kirk H, Trinh E, MacLeod T, et al. Complete mitochondrial genome of a gymnosperm, Sitka spruce (*Picea sitchensis*), indicates a complex physical structure. Genome Biol Evol. 2020;12:1174–9.

Jaramillo-Correa JP, Bousquet J. New evidence from mitochondrial DNA of a progenitor-derivative species relationship between black spruce and red spruce (Pinaceae). Am J Bot. 2003;90:1801–6.

Jaramillo-Correa JP, Bousquet J. Mitochondrial genome recombination in the zone of contact between two hybridizing conifers. Genetics. 2005;171:1951–62.

Jaramillo-Correa JP, Beaulieu J, Bousquet J. Contrasting evolutionary forces driving population structure at expressed sequence tag polymorphisms, allozymes and quantitative traits in white spruce. Mol Ecol. 2001;10:2729–40.

Jaramillo-Correa JP, Beaulieu J, Bousquet J. Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. Mol Ecol. 2004;13:2735–47.

Jaramillo-Correa JP, Beaulieu J, Ledig FT, Bousquet J. Decoupled mitochondrial and chloroplast DNA population structure reveals Holocene collapse and population isolation in a threatened Mexican-endemic conifer. Mol Ecol. 2006;15:2787–800.

Jaramillo-Correa JP, Beaulieu J, Khasa DP, Bousquet J. Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. Can J For Res. 2009;39:286–307.

Jaramillo-Correa JP, Bousquet J, Beaulieu J, Isabel N, Perron M, Bouillé M. Cross-species amplification of mitochondrial DNA sequence-tagged-site markers in conifers: the nature of polymorphism and variation within and among species in *Picea*. Theor Appl Genet. 2003;106:1353–67.

Jaramillo-Correa JP, Gérardi S, Beaulieu J, Ledig FT, Bousquet J. Inferring and outlining past population declines with linked microsatellites: a case study in two spruce species. Tree Genet Genomes. 2015;11:9.

Jeandroz S, Bastien D, Chandelier A, Du Jardin P, Favre JM. A set of primers for amplification of mitochondrial DNA in *Picea abies* and other conifer species. Mol Ecol Resour. 2002;2:389–92.

Jia ZR, Wang JH, Zhang SG. Pollen morphology and its phylogenetic implications in the genus *Picea*. Plant Syst Evol. 2014;300:461–73.

Johnsen KH, Major JE, Loo J, McPhee D. Negative heterosis not apparent in 22-year-old hybrids of *Picea mariana* and *Picea rubens*. Can J Bot. 1998;76:434–9.

Jokipii-Lukkari S, Sundell D, Nilsson O, Hvidsten TR, Street NR, Tuominen H. NorWood: a gene expression resource for evo-devo studies of conifer wood development. New Phytol. 2017;216:482–94.

Joost A, Bonin A, Bruford MW, Després L, Conord C, Erhardt G, et al. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol Ecol. 2007;16:3955–69.

Kang BY, Mann IK, Major JE, Rajora OP. Near-saturated and complete genetic linkage map of black spruce (*Picea mariana*). BMC Genomics. 2010;11:515.

Kang BY, Major JE, Rajora OP. A high-density genetic linkage map of a black spruce (*Picea mariana*) × red spruce (*Picea rubens*) interspecific hybrid. Genome. 2011;54:128–43.

Katsuki T, Shimada K, Yoshimaru H. Process to extinction and genetic structure of a threatened Japanese conifer species, *Picea koyamae*. J For Res. 2011;16:292–301.

Kim J-S, Kug J-S, Joeong S-J, Huntzinger DN, Michalak AM, Schwalm CR, et al. Reduced North American terrestrial primary productivity linked to anomalous Arctic warming. Nat Geosci. 2017;10:572–7.

Krutovskii KV, Bergman F. Introgressive hybridization and phylogenetic relationships between Norway, *Picea abies* (L.) Karst., and Siberian, *P. obovata* Ledeb., spruce species studied by isozyme loci. Heredity. 1995;74:464–80.

Kug J-S, Jeong J-H, Jang Y-S, Kim B-M, Folland CK, Min S-K, et al. Two distinct influences of Arctic warming on cold winters over North America and East Asia. Nat Geosci. 2015;8:759–63.

Lamara M, Raherison E, Lenz P, Beaulieu J, Bousquet J, MacKay J. Genetic architecture of wood properties based on association analysis and co-expression networks in white spruce. New Phytol. 2016;210:240–55.

Lamara M, Parent GJ, Giguère I, Beaulieu J, Bousquet J, MacKay JJ. Association genetics of acetophenone defence against spruce budworm in mature white spruce. BMC Plant Biol. 2018;18:231.

Laoué J, Depardieu C, Gérardi S, Lamothe M, Bomal C, Azaiez A, et al. Combining QTL mapping and transcriptomics to decipher the genetic architecture of phenolic compounds metabolism in the conifer white spruce. Front Plant Sci. 2021;112:793.

Le Corre V, Kremer A. The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol. 2012;21:1548–66.

Lebedev VG, Lebedeva TN, Chernodubov AI, Shestibratov KA. Genomic selection in forest tree improvement: methods, achievements and perspectives. Forests. 2020;11:1190.

Ledig FT, Jacob-Cervantes V, Hodgskiss PD, Eguiluz-Piedra T. Recent evolution and divergence among populations of a rare Mexican endemic, Chihuahua spruce, following Holocene climatic warming. Evolution. 1997;51:1815–27.

Ledig FT, Bermejo-Velázquez B, Hodgskiss PD, Johnson DR, Flores-López C, Jacob-Cervantes. The mating system and genetic diversity in Martínez spruce, an extremely rare endemic of México Sierre Madre Oriental: an example of facultative selfing and survival in interglacial refugia. Can J For Res. 2000a;30:1156–64.

Ledig FT, Mápula-Larreta M, Bermejo-Velázquez B, Reyes-Hernández V, Flores-López C, Capó-Arteaga MA. Locations of endangered spruce populations in México and the demography of *Picea chihuahuana*. Madroño. 2000b;47:71–88.

Ledig FT, Hodgskiss PD, Jacob-Cervantes V. Genetic diversity, mating system, and conservation of a Mexican subalpine relict, *Picea mexicana* Martínez. Conserv Genet. 2002;3:113–22.

Ledig FT, Hodgskiss P, Krutovskii KV, Neale DB, Eguiluz-Piedra T. Relationships among the spruces (*Picea*, Pinaceae) of Southwestern North America. Syst Bot. 2004;29:275–95.

Ledig FT, Hodgskiss P, Johnson DR. The structure of genetic diversity in Engelmann spruce and a comparison with blue spruce. Can J Bot. 2006;84:1806–28.

Ledig FT, Rehfeldt GE, Sáenz-Romero C, Flores-López C. Projections of suitable habitat for rare species under global warming scenarios. Am J Bot. 2010;97:970–87.

Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. J Dairy Sci. 2009;92:4656–63.

Lenz P, Beaulieu J, Mansfield SD, Clément S, Desponts M, Bousquet J. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of the boreal black spruce. BMC Genomics. 2017;18:335.

Lenz P, Nadeau S, Mottet M-J, Perron M, Isabel N, Beaulieu J. Multi-trait genomic selection for weevil resistance, growth and wood quality in Norway spruce. Evol Appl. 2020a;13:76–94.

Lenz P, Nadeau S, Azaiez A, Gérardi S, Deslauriers M, Perron M, et al. Genomic prediction for hastening and improving efficiency of forward selection in conifer polycross mating designs: an example from white spruce. Heredity. 2020b;124:562–78.

LePage BA. The evolution, biogeography and palaeoecology of the *Pinaceae* based on fossil and extant representatives. Acta Hortic. 2003;615:29–52.

Leslie AB, Beaulieu JM, Rai HS, Crane PR, Donoghue MJ, Mathews S. Hemisphere-scale differences in conifer evolutionary dynamics. Proc Natl Acad Sci U S A. 2012;109:16217–21.

Li Y, Stocks M, Hemmilä S, Källman T, Zhu H, Zhou Y, et al. Demographic histories of four spruce (*Picea*) species of the Qinghai-Tibetan Plateau and neighboring areas inferred from multiple nuclear loci. Mol Biol Evol. 2010;27:1001–14.

Li L, Abbot RJ, Liu B, Sun Y, Li L, Zou J, et al. Pliocene intraspecific divergence and Plio-Pleistocene range expansions within *Picea likiangensis* (Lijiang spruce), a dominant forest tree of the Qinghai-Tibet Plateau. Mol Ecol. 2013;22:5237–55.

Li L, Sun Y, Zou J, Yue W, Wang X, Liu J. Origin and speciation of *Picea schrenkiana* and *Picea smithiana* in the Center Asian Highlands and Himalayas. Plant Mol Biol Rep. 2015;33:661–72.

Li Z, De La Torre AR, Sterck L, Cánovas FM, Avila C, Merino I, et al. Single-copy genes as molecular markers for phylogenomic studies in seed plants. Genome Biol Evol. 2017;9:1130–47.

Li J, Xu B, Yang Q, Liu Z-L. The complete chloroplast genome sequence of *Picea schrenkiana* (Pinaceae). Mitochondrial DNA B Resour. 2020;5:2191–2.

Lin C-P, Huang J-P, Wu C-S, Hsu C-Y, Chaw S-M. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. Genome Biol Evol. 2010;2:504–17.

Lin D, Coombe L, Jackman S, Gagalova KK, Warren RL, Hammond SA, et al. Complete chloroplast genome sequence of an Engelmann spruce (*Picea engelmannii*) genotype from western Canada. Microbiol Resour Announc. 2019;8:e00382–19.

Lind M, Källman T, Chen J, Ma X-F, Bousquet J, Morgante M, et al. A *Picea abies* linkage map based on SNP markers identifies QTLs for four aspects of resistance to *Heterobasidion parviporum* infection. PLoS One. 2014;9:e101049.

Liu T-S. A new proposal for the classification of the genus *Picea*. Acta Phytotax Geobot. 1982;33:227–45.

Lo T, Coombe L, Lin D, Warren RL, Kirk H, Pandoh P, et al. Complete chloroplast genome sequence of a black spruce (*Picea mariana*) from Eastern Canada. Microbiol Resour Announc. 2020;9:e00877–20.

Lockwood JD, Aleksić JM, Zou J, Wang J, Liu J, Rennera SS. A new phylogeny for the genus *Picea* from plastid, mitochondrial, and nuclear sequences. Mol Phylogenet Evol. 2013;69:717–27.

Luikart G, Kardos M, Hand BK, Rajora OP, Aitken SN, Hohenlohe PA. Population genomics: advancing understanding of nature. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Springer; 2018. p. 3–79.

Manley SAM. Identification of red, black and hybrid spruces. Canadian Forestry Service Publication No 1301, Ottawa; 1971.

Meng L, Yang R, Abbott RJ, Miehe G, Hu T, Liu J. Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai-Tibetan Plateau and adjacent highlands. Mol Ecol. 2007;16:4128–37.

Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:1819–29.

Mikkola L. Observations on interspecific sterility in *Picea*. Ann Bot Fenn. 1969;6:285–339.
Milesi P, Berlin M, Chen J, Orsucci M, Li L, Jansson G, et al. Assessing the potential for assisted gene flow using past introduction of Norway spruce in southern Sweden: local adaptation and genetic basis of quantitative traits in trees. Evol Appl. 2019;12:1946–59.
Mimura M, Aitken SN. Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. Heredity. 2007;99:224–32.
Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci. 2009;92:4648–55.
Mukrimin M, Kovalchuk A, Neves LG, Jaber EHA, Haapanen M, Kirst M, et al. Genome-wide exon-capture approach identifies genetic variants of Norway spruce genes associated with susceptibility to *Heterobasidion parviporum* infection. Front Plant Sci. 2018;9:793.
Mullin TJ, Andersson B, Bastien J-C, Beaulieu J, Burdon RD, Dvorak WS, et al. Economic importance, breeding objectives and achievements. In: Plomion C, Bousquet J, Kole C, editors. Genetics, genomics and breeding of conifers. New York: CRC Press; 2011. p. 40–127.
Murray BG, Leitch IJ, Bennett MD. Gymnosperm DNA C-values database (release 5.0). 2012. http://www.data.kew.org/cvalues/. Accessed 18 Jan 2018
Nakvasina EN, Volkov AG, Prozherina NA. Provenance experiment with spruce (*Picea abies* (L.) Karst. and *Picea obovata* (Ledeb.)) in the North of Russia (Ahkhangelsk region). Folia Forestalia Polonica Ser A Forestry. 2017;59:219–30.
Namroud M-C, Beaulieu J, Juge N, Bousquet J. Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. Mol Ecol. 2008;17:3599–613.
Namroud M-C, Guillet-Claude C, MacKay J, Isabel N, Bousquet J. Molecular evolution of regulatory genes in spruces from different species and continents: heterogeneous patterns of linkage disequilibrium and selection but correlated recent demographic changes. J Mol Evol. 2010;70:371–86.
Namroud M-C, Bousquet J, Doerksen T, Beaulieu J. Scanning SNPs from a large set of expressed genes to assess the impact of artificial selection on the undomesticated genetic diversity of white spruce. Evol Appl. 2012;5:641–55.
Napier JD, de Lafontaine G, Hu FS. Exploring genomic variation associated with drought stress in *Picea mariana* populations. Ecol Evol. 2020;10:9271–82.
Nasri N, Bojovic S, Vendramin GG, Fady B. Population genetic structure of the relict Serbian spruce, *Picea omorika*, inferred from plastid DNA. Plant Syst Evol. 2008;271:1–7.
Neale DB, Savolainen O. Association genetics of complex traits in conifers. Trends Plant Sci. 2004;9:325–30.
Nystedt B, Street NR, Wetterborn A, Zuccolo A, Lin Y-C, Scofield DG, et al. The Norway spruce genome sequence and conifer genome evolution. Nature. 2013;497:579–84.
O'Connell LM, Mosseler A, Rajora OP. Extensive long-distance pollen dispersal in a fragmented landscape maintains genetic diversity in white spruce. J Hered. 2007;98:640–5.
Okuyama Y, Fujii N, Wakabayashi M, Kawakita A, Ito M, Watanabe M, et al. Nonuniform concerted evolution and chloroplast capture: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *mitella* (Saxifragaceae). Mol Biol Evol. 2004;22:285–96.
Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation adaptation and isolation by colonization. Mol Ecol. 2013;22:5983–99.
Otis Prud'homme G, Lamhamedi MS, Benomar L, Rainville A, De Blois J, Bousquet J, et al. Ecophysiology and performance of white spruce seed sources along a climatic gradient indicate the need for assisted migration. Front Plant Sci. 2018;8:2214.
Ouyang F, Hu J, Wang J, Ling J, Wang Z, Wang N, et al. Complete plastome sequences of *Picea asperata* and *P. crassifolia* and comparative analyses with *P. abies* and *P. morrisonicola*. Genome. 2019;62:317–28.

Parducci L, Jørgensen T, Tollefsrud MM, Elverland E, Alm T, Fontant SL, et al. Glacial survival of boreal trees in Northern Scandinavia. Science. 2012;335:1083–6.

Park Y-S, Beaulieu J, Bousquet J. Multi-varietal forestry integrating genomic selection and somatic embryogenesis. In: Park Y-S, Bonga JM, Moon H-K, editors. Vegetative progagation of forest trees. Seoul: National Institute of Forest Science; 2016. p. 302–22.

Pavy N, Parsons L, Paule C, Mackay J, Bousquet J. Automated SNP detection from a large collection of white spruce expressed sequences: contributing factors and approaches for the categorization of SNPs. BMC Genomics. 2006;7:174.

Pavy N, Pelgas B, Beauseigle S, Blais S, Gagnon F, Gosselin I, et al. Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. BMC Genomics. 2008;9:21.

Pavy N, Pelgas B, Laroche J, Rigault P, Isabel N, Bousquet J. A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. BMC Biol. 2012a;10:84.

Pavy N, Namroud MC, Gagnon F, Isabel N, Bousquet J. The heterogeneous levels of linkage disequilibrium in white spruce genes and comparative analysis with other conifers. Heredity. 2012b;108:273–84.

Pavy N, Gagnon F, Rigault P, Blais S, Deschênes A, Boyle B, et al. Development of high-density SNP genotyping arrays for white spruce (*Picea glauca*) and transferability to subtropical and nordic congeners. Mol Ecol Resour. 2013a;13:324–36.

Pavy N, Deschênes A, Blais S, Lavigne P, Beaulieu J, Isabel N, et al. The landscape of nucleotide polymorphism among 13,500 genes of the conifer *Picea glauca*, relationships with functions, and comparison with *Medicago truncatula*. Genome Biol Evol. 2013b;5:1910–25.

Pavy N, Gagnon F, Deschênes A, Boyle B, Beaulieu J, Bousquet J. Development of highly reliable *in silico* SNP resource and genotyping assay from exome capture and sequencing: an example from black spruce (*Picea mariana*). Mol Ecol Resour. 2016;16:558–98.

Pavy N, Lamothe M, Pelgas B, Gagnon F, Birol I, Bohlmann J, et al. A high-resolution reference genetic map positioning 8.8 K genes for the conifer white spruce: structural genomics implications and correspondence with physical distance. Plant J. 2017;90:189–203.

Payette S, Filion L. White spruce expansion at the tree line and recent climatic change. Can J Fort Res. 1985;15:241–51.

Peh KS-H, Corlett RT, Bergeron Y. Routledge handbook of forest ecology. New York: Routledge; 2015.

Pelgas B, Beauseigle S, Acheré V, Jeandroz S, Bousquet J, Isabel N. Comparative genome mapping among *Picea glauca*, *P. abies* and *P. mariana* x *rubens*, and correspondance with other Pinaceae. Theor Appl Genet. 2006;113:1371–93.

Pelgas B, Bousquet J, Meirmans PG, Ritland K, Isabel N. QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. BMG Genomics. 2011;12:145.

Perron M, Bousquet J. Natural hybridization between black spruce and red spruce. Mol Ecol. 1997;6:725–34.

Perron M, Gordon AG, Bousquet J. Species-specific RAPD fingerprintings for the closely related *Picea mariana* and *Picea rubens*. Theor Appl Genet. 1995;91:142–9.

Perron M, Perry DJ, Andalo C, Bousquet J. Evidence from sequence-tagged-site markers of a recent progenitor-derivative species pair in conifers. Proc Natl Acad Sci U S A. 2000;97:11331–6.

Perry DJ, Bousquet J. Genetic diversity and mating system of post-fire and post-harvest black spruce: an investigation using codominant sequence-tagged site (STS) markers. Can J For Res. 2001;31:32–40.

Petit RJ, Aguinagalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, et al. Glacial refugia: hotspots but not melting pots of genetic diversity. Science. 2003;300:1563–5.

Prunier J, Laroche J, Beaulieu J, Bousquet J. Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. Mol Ecol. 2011;20:1702–16.

Prunier J, Gérardi S, Laroche J, Beaulieu J, Bousquet J. Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. Mol Ecol. 2012;21:4270–86.

Prunier J, Pelgas B, Gagnon F, Desponts M, Isabel N, Beaulieu J, et al. The genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. BMC Genomics. 2013;14:368.

Prunier J, Tessier G, Bousquet J, MacKay J. From genotypes to phenotypes: expression levels of genes encompassing adaptive SNPs in black spruce. Plant Cell Rep. 2015;34:2111–5.

Prunier J, Caron S, MacKay J. CNVs into the wild: screening the genomes of conifer trees (*Picea* spp.) reveals fewer gene copy number variations in hybrids and links to adaptation. BMC Genomics. 2017a;18:97.

Prunier J, Caron S, Lamothe M, Blais S, Bousquet J, Isabel N, et al. Gene copy number variations in adaptive evolution: the genomic distribution of gene copy number variations revealed by genetic mapping and their adaptive role in an undomesticated species, white spruce (*Picea glauca*). Mol Ecol. 2017b;26:5989–6001.

Quiñones-Pérez Z, Simetal-Rodríguez SL, Sáenz-Romero C, Jaramillo-Correa JP, Wehenkel C. Spatial genetic structure in the very rare and species-rich *Picea chihuahuana* tree community (Mexico). Silvae Genet. 2014;63:149–59.

Raherison ES, Rigault P, Caron S, Poulin P-L, Boyle B, Verta J-P, et al. Transcriptome profiling in conifers and the PiceaGenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. BMC Genomics. 2012;13:434.

Raherison ES, Giguère I, Caron S, Lamara M, MacKay JJ. Modular organization of the white spruce (*Picea glauca*) transcriptome reveals functional organization and evolutionary signatures. New Phytol. 2015;207:172–87.

Rajora OP, Dancik B. Population genetic variation, structure, and evolution in Engelmann spruce, white spruce, and their natural hybrid complex in Alberta. Can J Bot. 2000;78:768–80.

Rajora OP, Eckert AJ, Zinck JWR. Single-locus versus multilocus patterns of local adaptation to climate in Eastern White Pine (*Pinus strobus*, Pinaceae). PLoS One. 2016;11:e0158691.

Ran J-H, Wei X-X, Wang X-Q. Molecular phylogeny and biogeography of *Picea* (Pinaceae): implications for phylogeographical studies using cytoplasmic haplotypes. Mol Phylogenet Evol. 2006;41:405–19.

Ran J-H, Shen T-T, Liu W-J, Wang P-P, Wang X-Q. Mitochondrial introgression and complex biogeographic history of the genus *Picea*. Mol Phylogenet Evol. 2015;93:63–76.

Ratcliffe B, Gamal El-Dien O, Klápště J, Porth I, Chen C, Jaquish B, et al. A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* × *glauca*) using unordered SNP imputation methods. Heredity. 2015;115:547–55.

Ratcliffe B, Gamal El-Dien O, Cappa EP, Porth I, Klápště J, Chen C, et al. Single-step BLUP with varying genotyping effort in open-pollinated *Picea glauca*. G3 Genes Genom Genet. 2017;7:935–42.

Rieseberg LH, Soltis DE. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol Trends Plants. 1991;5:65–84.

Rieseberg LH, Kim S-C, Randell RA, Whitney KD, Gross BL, Lexer C, et al. Hybridization and the colonization of novel habitats by annual sunflowers. Genetica. 2007;129:149–65.

Rigault P, Boyle B, Lepage P, Cooke JEK, Bousquet J, MacKay JJ. A white spruce gene catalog for conifer genome analyses. Plant Physiol. 2011;157:14–28.

Rodríguez-Quilón I, Santos-del-Blanco L, Serra-Varela MJ, Koskela J, González-Martínez SC, Alía R. Capturing neutral and adaptive genetic diversity for conservation in a highly structured tree species. Ecol Appl. 2016;26:2254–66.

Ru D, Mao K, Zhang L, Wang X, Lu Z, Sun Y. Genomic evidence for polyphyletic origins and interlineage gene flow within complex taxa: a case study of *Picea brachytyla* in the Qinghai-Tibet Plateau. Mol Ecol. 2016;25:2373–86.

Ru D, Sun Y, Wang D, Chen Y, Wang T, Hu Q, et al. Population genomic analysis reveals that homoploid hybrid speciation can be a lengthy process. Mol Ecol. 2018;27:4875–87.

Savard L, Li P, Strauss SH, Chase MW, Michaud M, Bousquet J. Chloroplast and nuclear gene sequences indicate Late Pennsylvanian time for the last common ancestor of extant seed plants. Proc Natl Acad Sci U S A. 1994;91:5163–7.

Scalfi M, Mosca E, Di Pierro EA, Troggio M, Vendramin GG, Sperisen C, et al. Micro- and macro-geographic scale effect on the molecular imprint of selection and adaptation in Norway spruce. PLoS One. 2014;9:e115499.

Scotti I, Burelli A, Cattonaro F, Chagné D, Fuller J, Hedley PE, et al. Analysis of the distribution of marker classes in a genetic linkage map: a case study in Norway spruce (*Picea abies* Karst.). Tree Genet Genomes. 2005;1:93–102.

Sigurgeirsson A, Szmidt AE. Phylogenetic and biogeographic implications of chloroplast DNA variation in *Picea*. Nord J Bot. 1993;13:233–46.

Sønstebø JH, Tollefsrud MM, Myking T, Steffenrem A, Nilsen AE, Edvardsen ØM, et al. Genetic diversity of Norway spruce (*Picea abies* (L.) Karst.) seed orchard crops: effects of number of parents, seed year, and pollen contamination. For Ecol Manage. 2018;411:132–141.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. Tree Genet Genomes. 2013;9:901–11.

Sperisen C, Büchler U, Gugerli F, Mátyás G, Geburek T, Vendramin GG. Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. Mol Ecol. 2001;10:257–63.

Stival Sena J, Giguère I, Rigault P, Bousquet J, MacKay J. Expansion of the dehydrin gene family in the Pinaceae is associated with considerable structural diversity and drought-responsive expression. Tree Physiol. 2018;38:442–56.

Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100:9440–5.

Storfer A, Patton A, Fraik AK. Navigating the interface between landscape genetics and landscape genomics. Front Genet. 2018;9:68.

Sullivan AR, Schiffthaler B, Thompson SL. Interspecific plastome recombination reflects ancient reticulate evolution in *Picea* (Pinaceae). Mol Biol Evol. 2017;34:1689–701.

Sullivan AR, Eldfjell Y, Schiffthaler B, Delhomme N, Asp T, Hebelstrup KH, et al. The mitogenome of Norway spruce and a reappraisal of mitochondrial recombination in plants. Genome Biol Evol. 2020;12:3586–98.

Sun Y, Abbot RJ, Li L, Li L, Zou J, Liu J. Evolutionary history of Purple cone spruce (*Picea purpurea*) in the Qinghai–Tibet Plateau: homoploid hybrid origin and Pleistocene expansion. Mol Ecol. 2014;23:343–59.

Sun Y, Abbott RJ, Lu Z, Mao K, Zhang L, Wang X, et al. Reticulate evolution within a spruce (*Picea*) species complex revealed by population genomic analysis. Evolution. 2018;72:2669–81.

Suren H, Hodgins KA, Yeaman S, Nurkowski KA, Smets P, Rieseberg LH, et al. Exome capture from the spruce and pine giga-genomes. Mol Ecol Resour. 2016;16:1136–46.

Sutton BCS, Flanagan DJ, Gawley JR, Newton CH, Lester DT, El-Kassaby YA. Inheritance of chloroplast and mitochondrial DNA in *Picea* and composition of hybrids from introgression zones. Theor Appl Genet. 1991;82:242–8.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F. Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol. 1998;7:453–64.

Taylor SA, Larson EL, Harrison RG. Hybrid zones: windows on climate change. Trends Ecol Evol. 2015;30:398–406.

Thistlethwaite FR, Gamal El-Dien O, Ratcliffe B, Klápště J, Porth I, Chen C, et al. Linkage disequilibrium vs. pedigree: genomic selection prediction accuracy in conifer species. PLoS One. 2020;15:e0232201.

Tollefsrud MM, Kissling R, Gugerli F, Johnsen Ø, Skrøppa T, Cheddadi R, et al. Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen. Mol Ecol. 2008;17:4134–50.

Tollefsrud MM, Latałowa M, van der Knaap WO, Brochmann C, Sperisen C. Late quaternary history of North Eurasian Norway spruce (*Picea abies*) and Siberian spruce *(Picea obovata*) inferred from macrofossils, pollen and cytoplasmic DNA variation. J Biogeogr. 2015;42:1431–42.

Trujillo-Moya C, George J-P, Fluch S, Geburek T, Grabner M, Karanitsch-Ackerl S, et al. Drought sensitivity of Norway spruce at the species' warmest fringe: quantitative and molecular analysis reveals high genetic variation among and within provenances. G3 Genes Genom Genet. 2018;8:1225–45.

Tsai C-J, Xue L-J. CRISPRing into the woods. GM Crops Food. 2015;6:206–15.

Tsuda Y, Chen J, Stocks M, Källman T, Sønstebø JH, Parducci L, et al. The extent and meaning of hybridization and introgression between Siberian spruce (*Picea obovata*) and Norway spruce (*Picea abies*): cryptic refugia as stepping stones to the west? Mol Ecol. 2016;25:2773–89.

Tulsieram LK, Glaubitz JC, Kiss G, Carlson JE. Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. Nat Biotechnol. 1992;10:686–90.

VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.

Vendramin GG, Anzidei M, Madaghiele A, Sperisen C, Bucci G. Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). Genome. 2000;43:68–78.

Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annu Rev Genet. 2013;47:97–120.

Wang X, Lu J, Yue W, Li L, Zou J, Li X, et al. Determining the extent and direction of introgression between three spruce species based on molecular markers from three genomes with different rates of gene flow. Plant Syst Evol. 2016;302:691–701.

Warren RL, Keeling CI, Yuen MMS, Raymond A, Taylor GA, Vandervalk BP, et al. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. Plant J. 2015;83:189–212.

Webb T III, Bartlein PJ. Global changes during the last 3 million years: climatic controls and biotic responses. Annu Rev Ecol Syst. 1992;23:141–73.

Wehenkel C, Saénz-Romero C. Estimating genetic erosion using the example of *Picea chihuahuana* Martínez. Tree Genet Genomes. 2012;8:1085–94.

Weng C, Jackson ST. Species differentiation of North American spruce (*Picea*) based on morphological and anatomical characteristics of needles. Can J Bot. 2000;78:1367–83.

Whitehill JGA, Yuen MMS, Bohlmann Y. Constitutive and insect-induced transcriptomes of weevil resistant and susceptible Sitka spruce. Plant-Environ Interact. 2021;2:137–47.

Wright JW. Species crossability in spruce in relation to distribution and taxonomy. For Sci. 1955;1:319–49.

Yang J-C, Joo M, So S, Yi D-K, Shin CH, Lee Y-M, et al. The complete plastid genome sequence of *Picea jezoensis* (Pinaceae: Piceoideae). Mitochondrial DNA Part A. 2016;27:3761–3.

Yeaman S, Whitlock MC. The genetic architecture of adaptation under migration–selection balance. Evolution. 2011;65:1897–911.

Yeaman S, Hodgins KA, Suren H, Nurkowski KA, Rieseberg LH, Holliday JA, et al. Conservation and divergence of gene expression plasticity following c. 140 million years of evolution in lodgepole pine (*Pinus contorta*) and interior spruce (*Picea glauca* × *Picea engelmannii*). New Phytol. 2014;203:578–91.

Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, et al. Convergent local adaptation to climate in distantly related conifers. Science. 2016;353:1431–3.

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38:203–8.

Zapata-Valenzuela J, Isik F, Maltecca C, Wegrzyn J, Neale DB, McKeand S, et al. SNP markers trace familial linkages in a cloned population of *Pinus taeda* – prospects for genomic selection. Tree Genet Genomes. 2012;8:1307–18.

Zhou L, Chen Z-Q, Olsson L, Grahn T, Karlsson B, Wu HX, et al. Effect of number of annual rings and tree ages on genomic predictive ability for solid wood properties of Norway spruce. BMC Genomics. 2020;21:323.

Zhu T, Hu J, Qi S, Ouyang F, Kong L, Wang J. Transcriptome and morpho-physiological analyses reveal factors regulating cone bud differentiation in Qinghai spruce (*Picea crassifolia* Kom.). Trees. 2021;35:1151–66.

Zou J, Sun Y, Li L, Wang G, Yue W, Lu Z, et al. Population genetic evidence for speciation pattern and gene flow between *Picea wilsonii*, *P. morrisonicola* and *P. neoveitchii*. Ann Bot London. 2013;112:1829–44.

Zou J, Yue W, Lili L, Wang X, Lu J, Duan B, et al. DNA barcoding of recently diversified tree species: a case study on spruces based on 20 DNA fragments from three different genomes. Trees. 2016;30:959–69.