

Resources for conifer functional genomics at the omics era

Rafael A. Cañas^{*,†}, M^a Belén Pascual, Fernando N. de la Torre, Concepción Ávila, Francisco M. Cánovas

Departamento de Biología Molecular y Bioquímica, Universidad de Málaga, Málaga, Spain

*Corresponding author: e-mail address: rcanas@uma.es

Contents

1. Introduction	2
2. Resources	4
2.1 Genome assemblies	4
2.2 Transcriptome assemblies	7
2.3 Expression microarrays	11
2.4 Online resources	16
2.5 Genetics for conifer functional genomics	19
3. Conclusions and perspectives	25
Acknowledgements	25
References	25

Abstract

Conifers are a group of woody plants that dominate vast forest ecosystems and represent a fundamental source of raw materials such as wood and other forest-derived products of great interest for humankind. Thus, understanding the molecular basis underlying the growth and development of these woody plants, as well as the response of these plants to environmental cues, is of paramount importance for the conservation of coniferous natural forests and the production of new tree lines with improved traits for sustainable forest plantations. Conifers have a series of characteristics that hinder the application of genetics and molecular biological approaches to generate data of high biological value. However, nowadays the biological sciences are dominated by omics techniques that have extensively increased the fundamental knowledge of conifer biology over the last decade. Novel resources are currently available for functional genomics research, and these resources are accessible to researchers all over the world. Recent advances in next-generation sequencing and assembly capacity have enabled whole-genome sequencing of several conifer species and have favoured the development of functional genomics studies in these non-model plant species. Currently, it is

[†] Senior author.

possible to obtain large amounts of data to explore the transcriptome or genome of any species with a limited budget. However, functional genomics studies using genetics approaches and gene manipulation remain tremendously challenging, and the application of recent gene editing techniques in conifers is awaited.



1. Introduction

Conifers are a group of woody gymnosperms that appeared during the Carboniferous period, approximately 300 million years ago (Farjon & Filer, 2013). This group of plants is distributed across all continents except Antarctica, dominating the forest ecosystems found in the vast boreal regions of North America and Eurasia (Farjon & Filer, 2013). These boreal forests are considered one of the major global biomes covering around the 11% of the land surface on Earth (Essen, Ehnström, Ericsson, & Sjöberg, 1997; Lorenz & Lal, 2010). Conifers are also an important source of industrial raw materials, including resins and wood (Farjon, 2010). Conifers accounted for the 64.8% of the total global production of sawlogs and veneer logs and for the 50.9% of that of pulpwood in 2015 (Fig. 1; FAOSTAT, <http://www.fao.org/faostat/en/?#home>).

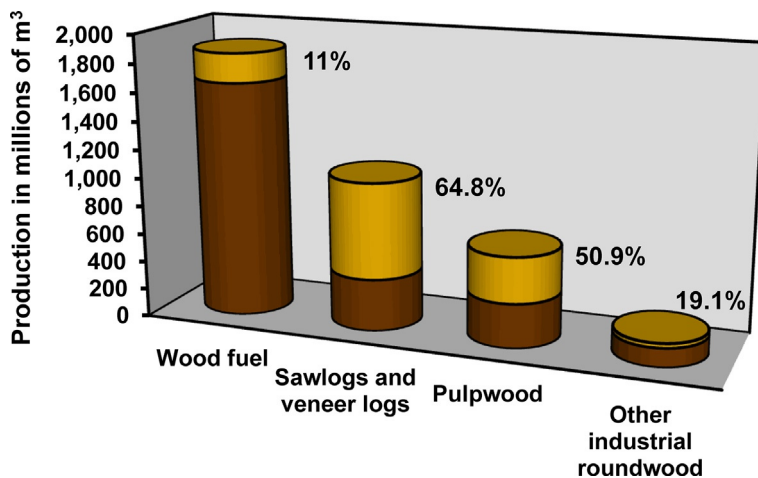


Fig. 1 World wood production in 2015 including the amount and percentage corresponding to conifers. Dark brown cylinders correspond to wood products from non-coniferous sources. Light brown cylinders correspond to wood products from coniferous sources. Data from FAOSTAT (<http://www.fao.org/faostat/en/>).

Over their evolutionary story, conifers have acquired a series of unique characteristics in response to interactions with their environments and to overcome biological challenges. There are very tall species, reaching up to 100m, such as Douglas fir (*Pseudotsuga menziesii*) (Domec et al., 2008). Most conifers are long-living evergreens, including species with extreme longevity such as *Pinus longaeva* (some specimens of which are approximately 5000 years old) (Farjon & Filer, 2013) and *Fitzroya cupressoides* (some specimens of which are approximately 3600 years old) (Lara & Villalba, 1993). Due to the long-life spans of conifers, these trees are exposed to changing environmental conditions, mainly due to seasonal changes (Cañas et al., 2015; Cañas, de la Torre, Pascual, Avila, & Cánovas, 2016). Therefore, some conifers species have developed extraordinary adaptations, such as the capacity of the Siberian *Picea obovata* to survive at extremely low temperatures (Kjellsen, Shiryaeva, Schröder, & Strimbeck, 2010).

Conifers also have very large genomes, which range from 20 to 40 Gb (De La Torre et al., 2014; Mackay et al., 2012; Ritland, 2012) and include a large number of repeated sequences with long telomeres up to 57 kb in size (Aronen & Ryyänänen, 2012; Flanary & Kletetschka, 2005; Neale et al., 2014), big introns (García-Gutiérrez, Cánovas, & Avila, 2018; Nystedt et al., 2013) and large amounts of transposable elements, mainly LTR (Long Terminal Repeat) retrotransposons, which may account for more than 70% of the genome (Cañas et al., 2017; Neale et al., 2014; Nystedt et al., 2013; Stevens et al., 2016).

In spite of the benefits of these biological characteristics of conifers, these characteristics have hindered conifer breeding, propagation and research, especially in the context of genetic manipulation (mutant libraries, production of transgenic plants, etc.). In some cases, these limitations have been overcome using large and costly infrastructures, which are available only to a small number of research groups. However, the emergence of “omics” techniques has partially palliated this problem at the research level. Of these techniques, next-generation sequencing (NGS) has allowed cheaper and faster sequencing of conifer genomes compared with classical reference models such as humans and several plant crop species, which would have been unbelievable in the recent past. Furthermore, these techniques have provided a way to explore gene function and perform functional genomics of conifers, including the characterization of the transcriptomes, proteomes and metabolomes but also the interactions and biological functions of their metabolic and regulatory pathways. An overview of recent

advancements is presented in this chapter, which attempts to trace the current status of the resources that are available for functional genomics of conifers.



2. Resources

Nearly every current omics technique requires the development of a reference dataset, such as genome and transcriptome sequences or metabolite libraries. An example of such a dataset is a physical or virtual cDNA library containing enough transcriptomic information to identify cDNAs and design and synthesize oligonucleotides for microarray construction. When omics techniques first started to be used, the procurement of these sequence pools was very time consuming and costly. However, this situation has changed with the emergence of NGS platforms, allowing the rapid sequencing of transcriptomes and even entire genomes, which remains a real challenge in conifers due to the enormous genomes of these species (De La Torre et al., 2014). In conifer research, resources for omics techniques have greatly increased over the last decade, which has led to the increased research activity associated with some functional genomics methods.

The resources available for conifer functional genomics include genome and transcriptome sequence assemblies, microarrays and bioinformatic tools. Most of these resources are publicly available through online databases, which can also be considered as resources since they usually provide additional tools for the analysis and visualization of biological data.

2.1 Genome assemblies

Despite the large sizes and repetitive nature of conifer genomes, five genome assemblies have been published to date (Table 1): *Picea abies* (Nystedt et al., 2013), *Picea glauca* (Birol et al., 2013), *Pinus taeda* (Zimin et al., 2014), *Pinus lambertiana* (Stevens et al., 2016) and the first draft of the *P. menziesii* genome (Neale et al., 2017). In addition, the genome assembly of *Gnetum montanum*, a gnetophyte, has been recently published (Wan et al., 2018). Current phylogenomic analyses indicate that gnetophytes belong to the same clade as conifers (Li et al., 2017; Wickett et al., 2014).

Overall, the size and complexity of conifer genomes lead to these assemblies having too many contigs with very partitioned drafts. This characteristic along with the presence of very large introns in a considerable number of genes has also limited the bioinformatic identification of many genes. With the exception of *P. abies*, the number of high confidence genes is very

Table 1 Conifer genome assemblies.
Sequenced genomes

Species	Reference of last version	Online database	Genes ^a	Year of first version
<i>Pseudotsuga menziesii</i>	Neale et al. (2017)	http://treegenesdb.org	35,880	2017
<i>Pinus lambertiana</i>	Stevens et al. (2016)	http://treegenesdb.org	13,936	2016
<i>Pinus taeda</i>	Zimin et al. (2017)	http://treegenesdb.org	9024	2014
<i>Picea glauca</i>	Warren et al. (2015)	http://congenie.org	14,462	2013
<i>Picea abies</i>	Nystedt et al. (2013)	http://congenie.org	26,359	2013

^aHigh confidence genes.

low in every assembly (Table 1). Each one of these limitations has led to new assembly strategies in order to obtain better draft genomes (Warren et al., 2015; Zimin et al., 2017). The use of sequencing platforms rendering long reads such as PacBio (<https://www.pacb.com/>) and Oxford Nanopore (<https://nanoporetech.com/>) in combination with Illumina could led to high-quality genome assemblies in conifers.

In functional genomics, genome assemblies are excellent tools to analyse different pools of data from NGS sequencing, which can include RNA-Seq and chromatin-derived samples (Alonso & Stepanova, 2015; Pevsner, 2015). However, there have been no reports of conifer genome assemblies being used for these purposes. In the case of RNA-Seq, the use of genome assemblies as a reference not only is useful to quantify gene expression but also can allow the identification of new transcripts derived from alternative splicing of genes (Steijger et al., 2013). Therefore, due to the limitations imposed by the quality of the constructed genome, reference transcriptomes and *de novo* transcriptome assemblies are much more adequate for RNA-Seq analyses (Steijger et al., 2013). In contrast, RNA-Seq data can also be used to aid the annotation of conifer genomes by identifying exons and introns from protein coding and non-coding genes.

For chromatin sequencing methods, read mapping with a genome assembly is essential for analysis. However, this process is dependent on the completeness and accuracy of the assembly. Fragmented and poorly annotated drafts are not adequate references for these analyses and lead to

bias in sequence coverage (Meyer & Liu, 2014). Another problem associated with conifer genomes is the number of repeated sequences, accounting for almost 80% of the genome in some species such as *P. lambertiana* (Stevens et al., 2016). This phenomenon can make it challenging to use these genomes as references for read mapping due to the introduction of bias in the coverage of certain regions, which, in the case of conifers, could be very substantial.

An important utility of genome sequences is that these sequences can be used for the implementation of genome editing tools such as TALENs (Boch, 2011) or CRISPR/Cas9 (Charpentier & Doudna, 2013; Feng et al., 2013). The application of these tools requires precise knowledge of the target sequence (see Section 2.5). However, considering the difficulties in using forward and reverse genetics for functional validation in conifers, the use of genome assemblies in the future as a reference for certain methods such as ChIP-Seq (Johnson, Mortazavi, Myers, & Wold, 2007) and DAP-Seq (O'Malley et al., 2016) is crucial for developing functional genomic research in conifers. The identification of epigenomic regulation sites and transcription factor-binding sites in chromatin will be of great utility for conifer research in the near future. Several studies have paved the way by characterizing transcription factors associated with plant development (Alvarez et al., 2018, 2015; Bravo et al., 2017; Carlsbecker et al., 2013; Hedman, Zhu, von Arnold, & Sohlberg, 2013; Niu et al., 2016; Rupps, Raschke, Rümmler, Linke, & Zoglauer, 2016; Uddenberg, Akhter, Ramachandran, Sundström, & Carlsbecker, 2015; Uddenberg et al., 2013; Valledor, Pascual, Meijón, Escandón, & Cañal, 2015; Zhu, Moschou, Alvarez, Sohlberg, & Von Arnold, 2014; Zhu, Moschou, Alvarez, Sohlberg, & Arnold, 2016) and those associated with lignification and regulation of secondary metabolism (Bedon et al., 2010; Bedon, Grima-Pettenati, & Mackay, 2007; Bomal et al., 2008, 2014; Craven-Bartle, Pascual, Cánovas, & Ávila, 2013; Pascual, Canovas, & Avila, 2015; Pascual et al., 2016, 2018; Rueda-López, García-Gutiérrez, Cánovas, & Ávila, 2013). The application of these types of techniques will be extremely helpful in the functional characterization of regulatory networks involved in the control of essential processes in conifers. The elucidation of these regulatory networks will provide essential biological data to be employed in the development of new tree lines with improved traits. However, the currently available methods are insufficient, and it is necessary to improve the assemblies and annotations to implement accurate genome-wide analyses.

2.2 Transcriptome assemblies

Whole transcriptomes are fundamental tools to design expression microarrays, analyse RNA-Seq experiments, identify peptides in proteomic studies using mass spectrometric analyses, annotate genome assemblies and identify genetic markers such as single nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs) (Dufva, 2009; Lowe, Shirley, Bleackley, Dolan, & Shafee, 2017; Pevsner, 2015; Romero-Rodríguez, Pascual, Valledor, & Jorrín-Novó, 2014). However, each tissue and cell type has its own transcriptome, which depends on development, metabolic state, environmental conditions, *etc.* (Yang & Kim, 2015). Therefore, the transcriptome is dynamic and changes with the forms (alternative splicing and chemical modifications of N bases) and amounts of individual transcripts (Ozsolak & Milos, 2011). Thus, the whole transcriptome of a species is the set of different transcriptomes expressed in each of the cell types and under different conditions.

This biological complexity leads to technical difficulties in obtaining complete transcriptomic data. In addition, before the emergence of NGS technologies, the procurement of transcriptomic data was very laborious and costly. The first resources for the generation of large-scale transcriptomic data were expressed sequence tag (EST) databases (Adams *et al.*, 1991). The need for physical libraries of cDNA in *Escherichia coli* and the individual Sanger sequencing of each clone limited the number of EST sequences and the assembly of transcriptomes (Lowe *et al.*, 2017). In addition, the experimental procedures favoured the identification of transcripts coding for proteins, while sequences belonging to non-coding RNA (ncRNA) were not analysed because these sequences were considered sequencing errors or junk. Different EST databases were generated for conifer species for gene discovery and annotation (Table 2), and in addition, a substantial number of studies in conifers have developed ESTs for genetic marker identification (*e.g.* Bérubé *et al.*, 2007; Perry & Bousquet, 1998; Rungis *et al.*, 2004). The objectives and laboratory resources involved are indicative of the variety of efforts in EST studies, with projects ranging from 282 sequences in *Pinus pinea* (Alonso *et al.*, 2007) to 206,875 sequences in *Picea* sp. (Ralph *et al.*, 2008). Due to the lack of genomic information about conifers caused by the enormous sizes of the genomes of these species, the main objective of most of these EST studies was the construction of gene catalogues. Therefore, the research efforts to develop EST data can be considered the pioneering functional genomic studies of conifers. Additionally,

Table 2 Representative published EST data in conifer species. Sequences obtained by Sanger's method.

Article	Species	Tissues	Reads	Unigenes	Contings	Singletons
Allona et al. (1998)	<i>Pinus taeda</i>	Differentiating compression wood and side wood	1097	ND	ND	ND
Ujino-Ihara et al. (2000)	<i>Cryptomeria japonica</i>	Inner bark	2231	1398	353	1045
Kirst et al. (2003)	<i>Pinus taeda</i>	Xylem-forming tissues	59,797	20,377	8070	12,307
Ujino-Ihara, Taguchi, Yoshimura, and Tsumura (2003)	<i>Cryptomeria japonica</i>	Male and female cones	1210	1173	ND	ND
Pavy et al. (2005)	<i>Picea glauca</i>	Different tissues and treatments	49,101	16,578	9354	7224
Asiegbu, Nahalkova, and Li (2005)	<i>Pinus sylvestris</i>	Infected roots	966	656	147	509
Lorenz et al. (2006)	<i>Pinus taeda</i>	Water stressed roots	12,918	6765	2006	4759
Cairney et al. (2006)	<i>Pinus taeda</i>	Zygotic embryos	68,721	12,154	5274	6880
Yakovlev, Fossdal, Johnsen, Junttila, and Skjøppa (2006)	<i>Picea abies</i>	Bud burst	1585	673	270	403
Lee, Chan, Wang, and Chu (2006)	<i>Taiwania cryptomerioides</i>	Inner bark	973	303	144	159
Alonso et al. (2007)	<i>Pinus pinea</i>	Cultured cotyledons	282	245	24	221

Ralph et al. (2008)	<i>Picea</i> sp. (<i>P. sitchensis</i> , <i>P. glauca</i> , <i>P. glauca- engelmannii</i>)	Different tissues and treatments	206,875	46,745	19,941	26,804
Futamura et al. (2008)	<i>Cryptomeria japonica</i>	Male strobili	36,011	10,463	7686	2777
Li, Wu, Dillon, and Southerton (2009)	<i>Pinus radiata</i>	Xylem-forming tissues	5952	3304	952	2352
Chancerel et al. (2011)	<i>Pinus pinaster</i>	Different tissues and treatments	31,678	13,730	4483	9247
Rigault et al. (2011) ^a	<i>Picea glauca</i>	Different tissues and treatments	146,616 ^b	27,720	ND	ND
Fernández-Pozo et al. (2011)	<i>Pinus pinaster</i>	Different tissues and treatments	19,160	10,843	2814	8029
Fernández-Pozo et al. (2011)	<i>Pinus sylvestris</i>	Different tissues and treatments	674	461	92	369
Villalobos et al. (2012)	<i>Pinus pinaster</i>	Differentiating compression wood and side wood	8429	ND	ND	ND
Mann, Wegrzyn, and Rajora (2013)	<i>Picea mariana</i>	Needles	4594	2731	497	2234

^aAdditional reads from NGS.

^bIn addition to EST from Pavy et al. (2005) and Ralph et al. (2008).

ND, Not determined.

the resulting EST data were used to design different cDNA expression microarrays for conifer species such as those developed for *P. taeda* (Van Zyl et al., 2002), *P. sitchensis* (Ralph et al., 2006) and *Pinus pinaster* (Villalobos et al., 2012) (see Section 2.3).

However, times have changed, and NGS allows rapid and inexpensive procurement of transcriptomic sequences in comparison with the old approaches. NGS makes it possible for individual research groups working in non-model species to generate appropriate transcriptomic resources. Although these data could be used to design new expression microarrays, this approach has become meaningless since most transcriptomic data are obtained from RNA-Seq experiments, accounting for transcript accumulation. Thus, the main uses of transcriptomic data are as follows: (i) as a reference to map sequenced reads in RNA-Seq experiments (e.g. Cañas et al., 2015), (ii) as a guide for genome annotations (e.g. Warren et al., 2015; Wegrzyn et al., 2014), (iii) to generate a virtual proteome for peptide identification in proteomic experiments (e.g. Romero-Rodríguez et al., 2014), (iv) as a reference for the design of gene/exome capture experiments (Lu et al., 2016; Neves, Davis, Barbazuk, & Kirst, 2013, 2014; Seoane-Zonjic et al., 2016; Suren et al., 2016) and (v) as a source of information about genetic markers such as microsatellites or SSRs (e.g. Canales et al., 2014; Hu et al., 2016) (Fig. 2).

In most cases, the results obtained from a single experiment are sufficient for the analyses (e.g. RNA-Seq). However, in some situations, complex transcriptomic assemblies that use data from a variety of tissues or conditions

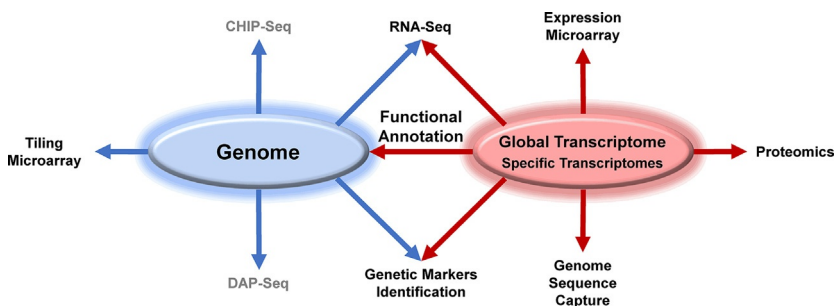


Fig. 2 Some of the main roles of genome and transcriptome assemblies as references in research. Techniques highlighted with blue lines are dependent on genome assemblies or sequences. Techniques highlighted with red lines are dependent on transcriptome assemblies or sequences. Techniques in grey still have not been developed in conifers due to fragmented genome assemblies. *ChIP-Seq*, chromatin immunoprecipitation sequencing; *DAP-Seq*, DNA affinity purification sequencing.

are required, for example, when using the transcriptome for genome annotation. In such cases, the reference transcriptome is used in an attempt to cover most of the transcripts expressed in a species, including splice variants and non-coding protein sequences (e.g. Canales et al., 2014; Lorenz et al., 2012; Rigault et al., 2011; Yeaman et al., 2014). In this regard, studies on microRNA (miRNA) are very limited, with some reports identifying and exploring miRNA expression under different physiological and developmental conditions (de Vega-Bartol et al., 2013; Fossdal et al., 2012; Liu & El-Kassaby, 2017; Morin et al., 2008; Niu et al., 2015; Qiu et al., 2015; Quinn, Iriyama, & Fernando, 2014; Wan et al., 2012; Xie et al., 2017; Zhang et al., 2013) or performing computational analyses in species with a genome draft, such as *P. taeda* (Quinn, Iriyama, & Fernando, 2015) and *P. glauca* (Yakovlev & Fossdal, 2017). Studies on long non-coding RNA (lncRNA) are nearly non-existent with the exception of the lncRNA catalogue for Norway spruce which was included in the first whole-genome assembly (Nystedt et al., 2013). Nevertheless, considering that a good genome assembly is necessary for classification of the identified lncRNA, it is understandable that lncRNAs have been poorly characterized in conifer species.

2.3 Expression microarrays

Expression microarray technology makes it possible to experimentally explore important portions of a transcriptome including miRNA by using specialized platforms. Expression microarrays are a hybridization-based technique that uses thousands of probes, cDNAs or oligonucleotides, corresponding to different transcripts fixed on a solid surface and hybridized with experimental samples labelled with fluorophores (the current state of the art is reviewed in Li, Sedighi, & Wang, 2016). In comparison with previous methods of gene expression analysis such as dot, slot and northern blots, the emergence of microarrays was an extraordinary advancement and provided the ability to evaluate the expression dynamics of thousands of transcripts simultaneously. Regarding expression studies in conifers, this methodology was used to complement and, finally, to overcome the transcript catalogue age of EST libraries. A significant, early study was published by Whetten, Sun, Zhang, and Sederoff (2001). Currently, in the NCBI's Gene Expression Omnibus (NCBI's GEO, <https://www.ncbi.nlm.nih.gov/geo/>) more than 30 microarray platforms for conifer species have been deposited since 2004 (Tables 3 and 4). This resource includes improved

Table 3 Gene expression microarrays available for conifers (*Pinidae* term in a platform search) at NCBI Gene Expression Omnibus (GEO).

GEO accession	Species	Type	Probe	Spot number	Year
No submitted ^a	<i>Pinus tabuliformis</i>	Commercial	<i>In situ</i> oligonucleotide	60,000	2016
GPL21977	<i>Pinus taeda</i>	Commercial	<i>In situ</i> oligonucleotide	180,880	2016
GPL22783	<i>Picea glauca</i>	Commercial	<i>In situ</i> oligonucleotide	177,321	2016
GPL19600	<i>Cryptomeria japonica</i>	Commercial	<i>In situ</i> oligonucleotide	71,984	2016
GPL21757	<i>Pinus sylvestris</i> ; <i>Cenococcum geophilum</i>	Commercial	<i>In situ</i> oligonucleotide	43,926	2016
GPL19234	<i>Pinus pinaster</i>	Commercial	<i>In situ</i> oligonucleotide	42,500	2015
GPL19078	<i>Pinus sylvestris</i>	Commercial	<i>In situ</i> oligonucleotide	36,349	2015
GPL18419	<i>Pinus pinaster</i>	Non-commercial	Spotted cDNA	8208	2015
GPL18152	<i>Cryptomeria japonica</i>	Commercial	<i>In situ</i> oligonucleotide	62,892	2014
GPL18030	<i>Cryptomeria japonica</i>	Commercial	<i>In situ</i> oligonucleotide	75,191	2013
GPL18004	<i>Pinus massoniana</i>	Commercial	Spotted oligonucleotide	48,598	2013
GPL14845	<i>Cryptomeria japonica</i>	Commercial	<i>In situ</i> oligonucleotide	45,753	2013
GPL16286	<i>Pinus taeda</i>	Commercial	<i>In situ</i> oligonucleotide	180,880	2012
GPL15915	<i>Pinus taeda</i>	Non-commercial	Spotted cDNA	26,496	2012
GPL15526	<i>Pinus pinaster</i>	Non-commercial	Spotted cDNA	4223	2012
GPL15510	<i>Pinus pinaster</i>	Non-commercial	Spotted cDNA	2910	2012
GPL15033	<i>Picea glauca</i> ; <i>Picea sitchensis</i>	Non-commercial	Spotted oligonucleotide	34,992	2011
GPL11184	<i>Pinus taeda</i>	Non-commercial	Spotted cDNA	26,496	2011

Table 3 Gene expression microarrays available for conifers (*Pinidae* term in a platform search) at NCBI Gene Expression Omnibus (GEO).—
cont'd

GEO accession	Species	Type	Probe	Spot number	Year
GPL10754	<i>Picea glauca</i>	Non-commercial	Spotted cDNA	11,036	2010
GPL8302	<i>Pinus radiata</i>	Non-commercial	Spotted cDNA	18,432	2009
GPL6582	<i>Pinus pinaster</i>	Commercial	Spotted cDNA	24,288	2008
GPL5189	<i>Pinus taeda</i>	Non-commercial	Spotted cDNA	49,152	2007
GPL5423	<i>Picea</i>	Non-commercial	Spotted cDNA	25,344	2007
GPL5763	<i>Pinus taeda</i>	Non-commercial	Spotted cDNA	5376	2007
GPL2027	<i>Pinus taeda</i> ; <i>Pinus elliottii</i> var. <i>elliottii</i>	Non-commercial	Spotted cDNA	8448	2006
GPL4039	<i>Pinus taeda</i>	Non-commercial	Spotted cDNA	6144	2006
GPL4657	<i>Pinus taeda</i>	Non-commercial	Spotted cDNA	5376	2006
GPL1880	<i>Pinus taeda</i>	Non-commercial	Spotted cDNA	4794	2005
GPL1014	<i>Pinus taeda</i>	Non-commercial	Spotted cDNA	9216	2004
GPL1508	<i>Pinus taeda</i>	Non-commercial	Spotted cDNA	9204	2004
No submitted ^b	<i>Pinus taeda</i>	Non-commercial	Spotted cDNA	3000	2002

^aPresented at [Niu et al. 2016](#).

^bPresented at [Whetten et al. \(2001\)](#).

Table 4 miRNA expression microarrays available for plants including conifers (*Pinidae* term in a platform search) at NCBI Gene Expression Omnibus (GEO).

GEO accession	Species	Type	Probe	Spot number	Year
GPL17805	Different plant species including <i>P. glauca</i> , <i>P. sitchensis</i> and <i>P. taeda</i>	Commercial	<i>In situ</i> oligonucleotide	1512	2013
GPL9858	Different plant species including <i>P. taeda</i>	Commercial	<i>In situ</i> oligonucleotide	3968	2011
GPL9344	Different plant species including <i>P. taeda</i>	Commercial	<i>In situ</i> oligonucleotide	3968	2009
GPL9736	Different plant species including <i>P. taeda</i>	Commercial	<i>In situ</i> oligonucleotide	3540	2009
GPL9738	Different plant species including <i>P. taeda</i>	Commercial	<i>In situ</i> oligonucleotide	3281	2009
GPL9737	Different plant species including <i>P. taeda</i>	Commercial	<i>In situ</i> oligonucleotide	3122	2009

versions of the first microarrays for determined species and some versions that cover plant miRNA catalogues with sequences from certain conifer species.

However, the nature and characteristics of this technology had led to major limitations in the use of microarrays in research compared with RNA-Seq. Some of the main problems associated with microarray technology in comparison to RNA-Seq are lower sensitivity, lower dynamic range, hybridization bias and that microarray technology only yields information about known transcripts (Martin, Dehler, & Król, 2016). Therefore, the depth of the analysis depends on the number and the origin of the probes deposited over the microarray. In contrast, in RNA-Seq, the depth of the analysis depends on the number of sequenced reads in the experiment. Thus, in expression microarrays, it is not possible to cover both gene expression and the expression of alternative transcript forms that are not physically contained in the microarray. Therefore, the previous work to design and produce the microarrays determines the final quality of the data obtained. Microarrays containing spotted cDNAs as probes can be limited by the number of analysed transcripts and by transcriptome coverage (Martin et al., 2016). cDNAs must be cloned and physically preserved but also arranged in plates to be printed on the microarray surface, which is time consuming, costly and requires a lot of space. In this context, covering the whole

transcriptome of a species can be a challenging objective. The transcriptome coverage of a microarray is, of course, strongly associated with the origins of the sequences (tissues, environmental, physiological and metabolic conditions). However, microarrays using synthesized oligonucleotides as probes allow more extensive coverage of the transcriptome. In this case, the sequences employed for oligonucleotide design come from the large transcriptomic databases that have become available since the advent of NGS (e.g. Cañas, Feito, et al., 2015; Raherison et al., 2012). The generation of microarrays has become more accessible making it possible to include a greater number of probes. In conifers, the platform with the higher number of spotted cDNA is GPL5189 for *P. taeda* with 49,152 probes (Table 3), while the oligonucleotide platform for a conifer species with the most probes has 180,880 probes (GPL21977 for *P. taeda*, Table 3), which is 3.7 times higher than the value obtained for GPL5189. Since 2012, concurrent with the rise of NGS, *in situ* oligonucleotide microarrays have been preferred to cDNA microarrays and are usually commercially manufactured by specialized companies such as Agilent or Affymetrix. NGS and RNA-Seq are increasingly replacing gene expression microarrays, although microarrays have demonstrated extraordinary utility over the last 15 years. Since 2012, there have been an increasing number of studies on conifers using RNA-Seq, and in the last few years, the number of studies using RNA-Seq has surpassed the number of publications using microarrays (Fig. 3). Currently, there is

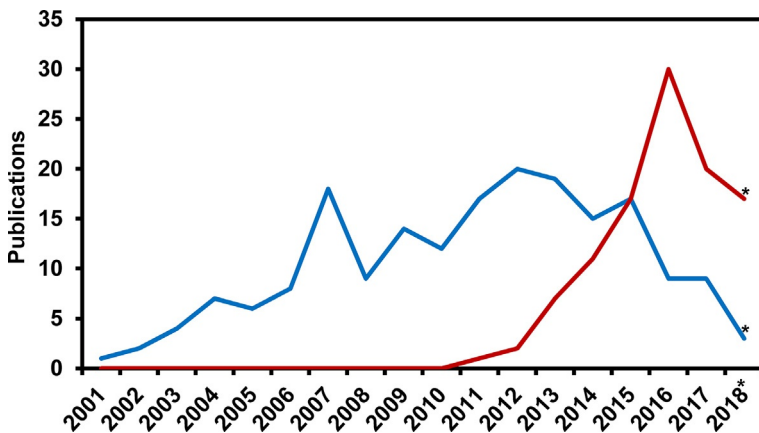


Fig. 3 Publications in conifers using microarrays and RNA-Seq. The publication numbers were estimated through a search in the Web of Science. Blue line shows microarray publications per year (search: “microarray AND conifer*”). Red line shows RNA-Seq publications per year (search: “RNA-Seq AND conifer*”). *Data accessed at 28th November 2018.

a clear trend towards performing transcriptomic analyses *via* NGS, which is mainly because the output of a large amount of data does not require previously determined information about the transcriptome. Each experiment is an opportunity to discover new transcript variants, including those with low expression levels. In the case of non-conventional model species such as conifers, the use of RNA-Seq is extremely advantageous because the transcriptome is obtained primarily from the assembly of sequencing data (see [Section 2.2](#)). Additionally, microarray hybridization can also lead to more false positives than the RNA-Seq procedures ([Cañas, Feito, et al., 2015](#)), which is mainly due to hybridization bias; the probe can hybridize with transcripts of different genes with extremely high identity (*e.g.* two paralog genes appearing in a recent genetic duplication) or does not differentiate between different transcripts of a single gene, mainly if spotted cDNA microarrays are used.

2.4 Online resources

New high-throughput techniques coupled with functional genomics produce large amounts of data requiring laborious analyses, resulting in most of the data remaining unseen. Therefore, access to data is an important issue in the omics era. The need to revisit the data and make new comparisons or visualizations has promoted the emergence of accessible online databases ([Galperin, Fernández-Suárez, & Rigden, 2017](#)) and web-servers ([Benson, 2017](#)). There are general web resources that can be employed for every species in biological research, for example, the three main databases containing nucleotide and protein sequences: NCBI's GenBank ([Benson et al., 2017](#)), DDBJ ([Mashima et al., 2016](#)) and ENA ([Gibson et al., 2016](#)). Whole-genome sequence assemblies of four conifer species have been deposited in NCBI's GenBank: *P. menziesii* ([GCA_001517045.1](#)); *P. glauca* ([GCA_000411955.5](#), [GCA_000966675.1](#) and [GCA_001687225.1](#)); *P. taeda* ([GCA_000404065.3](#)) and *P. lambertiana* ([GCA_001447015.2](#)). Although sequence information is basic, such resources may also store other types of functional genomic data and tools such as gene expression data from RNA-Seq and microarrays (GEO, [Barrett et al., 2013](#)), data regarding plant metabolic and regulatory pathways (Plant Reactome, [Naithani et al., 2017](#)) and mass-spectrometric data from proteomic analyses (ProteomeXchange, [Vizcaino et al., 2014](#)).

In addition to these general web resources, there are several databases and web-servers containing specific information and tools that can be used for functional genomic studies of conifers ([Table 5](#)). Usually, these resources

Table 5 Online resources for conifer functional genomics.

Name	Data type	Species	Link
Dendrome/ TreeGenes	Genome assemblies, transcriptomes	<i>P. taeda</i> , <i>P. glauca</i> , <i>P. sitchensis</i> , and other species	https://dendrome.ucdavis.edu/index.php/ http://tgwebprod.cam.uchc.edu/Drupal/
PineRefSeq	Genome assemblies	<i>P. taeda</i> , <i>P. lambertiana</i> , <i>P. menziesii</i>	http://pinegenome.org/pinerefseq/
ConGenIE	Genome assemblies, transcriptomes, gene expression, visualization and analysis tools	<i>P. abies</i> , <i>P. glauca</i> , <i>P. taeda</i> and other species	http://congenie.org/
NorWood	Wood gene expression network	<i>P. abies</i>	http://norwood.congenie.org/norwood-v1.0/
GymnoPLAZA	Transcriptomes	<i>P. taeda</i> , <i>P. abies</i> and other conifers	https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/
EVOLTREE	Genomic and EST sequences	<i>A. alba</i> , <i>P. abies</i> , <i>P. pinaster</i> , <i>P. sylvestris</i>	http://www.evoltree.eu/index.php
PineDB	Transcriptome	<i>P. taeda</i>	http://bioinfolab.muohio.edu/txid3352v1/index.php
Sugi Genome DB	Genome sequences Transcriptomes	<i>C. japonica</i> , <i>C. obtusa</i>	http://www.ffpri.affrc.go.jp/labs/cjgenome/index.html
SustainPineDB	Transcriptome	<i>P. pinaster</i>	http://www.scbi.uma.es/sustainpinedb/home_page

have been developed in the framework of research projects, and therefore most of these resources have either disappeared or have been integrated into other specific or general databases upon culmination of the project (e.g. Fernández-Pozo et al., 2011; Lippert, Yuen, & Bohlmann, 2009; Pavy et al., 2007). Among current online resources, Dendrome/TreeGenes is one of the older resources available (Wegrzyn, Lee, Tearse, & Neale, 2008). Dendrome/TreeGenes was developed under the framework of the Dendrome Project as a resource to integrate tools and tree-genomic data including data regarding species, publications, genetic maps, molecular markers, and genotypes as well as phenotypic data, genomic sequences, transcriptomes and expression data. Although the database was not conceived only for conifers, they constitute a fundamental part of the database. Of the 271 tree species in the transcriptome database 176 correspond to conifers; however, a substantial number of these sequences correspond to only a few species such as *P. taeda* and *P. sitchensis*. Another interesting web resource is PineRefSeq, which is closely associated with Dendrome/TreeGenes in terms of the project partners, primarily the University of California, Davis (UC Davis). Both databases provide FTP access to the genome assemblies and transcriptomes of three different conifers from North America (*P. taeda*, *P. menziesii* and *P. lambertiana*) including new unpublished versions.

Similar to Dendrome/TreeGenes, Conifer Genome Integrative Explorer (ConGenIE) is a web resource that integrates data about different conifer species, although this resource is mainly focused on *P. abies* as a part of the Norway spruce genome sequencing project (Sundell et al., 2015). This resource includes the genome assemblies of *P. abies*, *P. glauca* and *P. taeda*; transcriptomic sequences of these three species and different conifer and gymnosperm species; gene expression data from *P. abies*; and different visualization and analytic tools. ConGenIE is part of the Plant Genome Integrative Explorer (PlantGenIE) platform, which includes genomic resources for *Populus* (PopGenIE) and *Arabidopsis* (AtGenIE) (Sundell et al., 2015). An important characteristic of ConGenIE is the integration of this resource with additional web tools to analyse the gene expression data from *P. abies* or even other conifers such as *P. pinaster*. The Comparative Plant Expression tool (ComPIEx, <http://complex.plantgenie.org/>) allows the comparison and visualization of the gene co-expression networks of the three species in PlantGenIE (Netotea, Sundell, Street, & Hvidsten, 2014). Another example is the exImage tool for visualizing gene expression over organ and tissue pictograms in *P. abies* and more recently in *P. pinaster* (Cañas et al., 2017; <http://v22.popgenie.org/microdissection/>). A web resource associated with

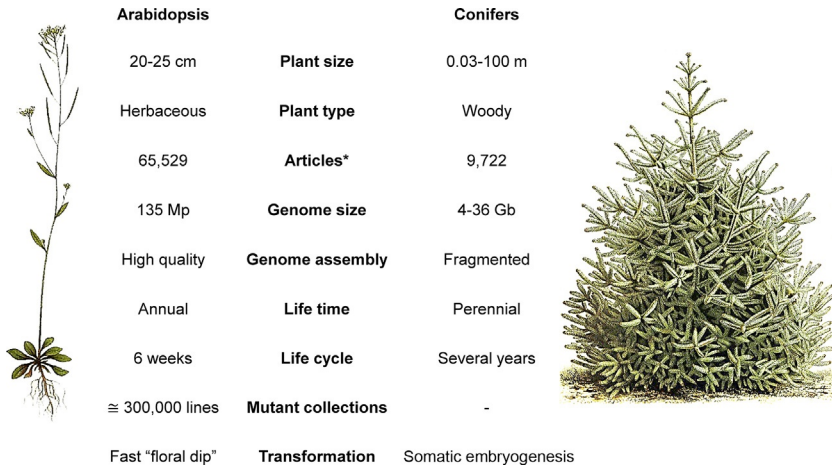
ConGenIE is NorWood, which contains gene expression profiles and co-expression networks from high-spatial-resolution tissue sections of the cambial and woody tissues of *P. abies* (Jokipii-Lukkari et al., 2017). Integration of the data from NorWood with PlantGenIE data from *Arabidopsis* and *Populus* provides evolutionary insight into plant wood formation.

Another web resource that includes data from different species is GymnoPLAZA, which is part of a bigger platform for plants, PLAZA (Proost et al., 2015). GymnoPLAZA contains sequence and gene annotation data from different conifer species and other gymnosperms and angiosperms. This web resource also provides tools for comparative genomics. GymnoPLAZA is a good source of data for functional genomics, although the main aim of this resource is the facilitation of comparative and evolutionary genomics. Other resources that were not developed for functional genomics but contain useful information for such studies in conifers are the transcriptomic and genomic databases generated in the framework of the EVOLTREE project (<http://www.evoltree.eu/>). These databases contain data from conifer species such as *Abies alba*, *P. abies*, *P. pinaster*, *P. sylvestris* and *Podocarpus oleifolius* and also from angiosperm species of the genera *Fagus* and *Quercus*.

Alternatively, several web resources containing data about specific conifer species have been mainly developed to store the results generated by other projects and research groups. For example, PineDB at Miami University (Oxford, OH, USA) is a web database that contains extensive transcriptomic information for *P. taeda* (Table 5). Among this class of resources, a constant and meritorious effort is that of the Sugi Genome DB, which centralizes the genetic, genomic and transcriptomic data of the Asian cypresses; *Cryptomeria japonica* (sugi, Japanese cedar) and *Chamaecyparis obtusa* (hinoki, Japanese cypress) (Table 5). This resource contains information that has been generated since 1998 by the Japanese Forestry and Forest Products Research Institute. Finally, the SustainPineDB contains the transcriptomic sequence information of the western Mediterranean conifer *P. pinaster* (Canales et al., 2014). This database replaced an earlier resource, the EuroPineDB (Fernández-Pozo et al., 2011), and was developed in the framework of the European research project SustainPine (<http://www.scbi.uma.es/sustainpine/>).

2.5 Genetics for conifer functional genomics

Genetic techniques are powerful tools for elucidating gene functions. In model plant species such as *Arabidopsis thaliana*, these techniques have been widely used for the functional characterization of genes by taking advantage



Arabidopsis		Conifers
20-25 cm	Plant size	0.03-100 m
Herbaceous	Plant type	Woody
65,529	Articles*	9,722
135 Mp	Genome size	4-36 Gb
High quality	Genome assembly	Fragmented
Annual	Life time	Perennial
6 weeks	Life cycle	Several years
≅ 300,000 lines	Mutant collections	-
Fast "floral dip"	Transformation	Somatic embryogenesis

Fig. 4 Characteristics of *Arabidopsis* vs. conifers. *Articles in the Web of Science from 2008 to 2017 (searches "*Arabidopsis thaliana*" and "Conifer").

of the biological characteristics of the model, including small plant and genome sizes, short lifespans and ease of transformation (Fig. 4). These advantages are complemented by the high number of labs and researchers working on this model, leading to a high potential for knowledge accumulation and to the development of new methods, mainly using reverse genetics, to study gene function (Provart et al., 2016). However, conifer species exhibit characteristics that make it challenging to apply certain genetic methods: large body sizes and long-life cycles (Farjon, 2010), large genome sizes (Ritland, 2012), high heterozygosity (Badenes, Fernández i Martí, Ríos, & Rubio-Cabetas, 2016) and difficulties in obtaining viable transgenic plants (Klimaszewska, Lachance, Pelletier, Lelu, & Seguin, 2001) (Fig. 4).

In this context, the genetic approaches used with conifers have been focussed on genotype-phenotype association, breeding, diversity and population structure more than on large-scale programs of mutant characterization or genetic engineering associated with functional genomics (Badenes et al., 2016; Plomion et al., 2016). The final aim of these studies is to assist and improve conifer breeding programs. Breeding programs for trees, including conifers, are large, expensive and long, requiring many years and resources to produce desired results (West, 2014). Many of these programs were started in the 1950s and have continued till the present (Pâques, 2013). Although these approaches are so powerful that important gains in the desired traits can be obtained in a few generations (West, 2014), there is a need in to accelerate these programs due to the long breeding cycles.

An alternative method to reach that goal is to use marker-assisted selection (MAS) of trees and quantitative trait loci (QTL) mapping. In recent years, the emergence of omics techniques has improved the identification and use of markers, mainly single nucleotide polymorphisms (SNPs), for gene mapping and genotyping. Transcriptome sequencing is widely employed, even in species rarely used in research, to identify thousands of SNPs (e.g. Liu et al., 2016; Wachowiak, Trivedi, Perry, & Cavers, 2015). The extent of genotyping has widened since the advent of SNP microarrays (e.g. Chancerel et al., 2013; Pavy et al., 2013, 2017), CGH microarrays for the identification of DNA copy number variations (CNVs) (Prunier, Caron, & MacKay, 2017) and exome capture sequencing, which also allows the identification of new markers (e.g. Lu et al., 2016; Neves et al., 2014; Pavy et al., 2016; Suren et al., 2016). Although SNPs in allelic variants can be considered mutations, most SNPs have no evident associations with a defined phenotype. In some cases, these relationships have been established allowing assignments of function to specific genes (e.g. Lu et al., 2017). A promising step forward in tree domestication seems to be the combined use of genomic selection (GS) of conifers with clonal propagation *via* somatic embryogenesis (Harfouche et al., 2012). This approach can substantially accelerate breeding programs. Unlike MAS, the fundamental principle of GS is the development of a prediction model generated from all available genetic markers and phenotypic data collected from a sample population. The aim of this model is to determine the genomic breeding value of progeny (Harfouche et al., 2012).

The use of genetics to search for or validate gene function is very limited in conifers. There are no mutant collections and extensive transformation programs are slow and expensive. In the case of mutant collections, screening for individual conifers with modified characteristics would be challenging due to their long-life cycles and large sizes of these trees, and the time and infrastructure required to complete the study would be tremendous. Thus, few studies have been carried out with mutant conifers, and usually, these studies have been conducted on trees with apparent phenotypes (Carlsbecker et al., 2013; Guo, Zheng, & Yang, 2014; Han, Shinohara, Kakubari, & Mukai, 2003; Hirao, Watanabe, Kurita, Kondo, & Takata, 2009; Ning, Lu, Liu, Yang, & Wang, 2013; Tsubomura, Kurita, & Watanabe, 2016; Uddenberg et al., 2013), although the best characterized conifer mutant, a mutant of the gene encoding the cinnamyl alcohol dehydrogenase enzyme (CAD; EC 1.1.1.195) in *P. taeda*, was identified in a genetic screen (MacKay et al., 1997; Ralph et al., 1997). NGS and other

omics techniques represent a new opportunity to characterize conifer mutants and, in some cases, to identify the mutated gene. An excellent example of such a study is that performed on the Norway spruce mutant *acrocona*, in which a MADS box transcription factor was identified as the regulator of early cone setting using RNA-Seq (Uddenberg et al., 2013).

Gene function assignment by using spontaneous mutants can also be driven by gene silencing using RNA interference (RNAi) (Watanabe, 2011). Despite the challenges associated with transforming conifers, gene knockdown characterization has provided the data required to explore candidate genes and assign molecular and physiological functions to these genes (e.g. Maloney, Samuels, & Mansfield, 2012; Moschou et al., 2016; Zhu et al., 2014). Similar to the study of mutants, the combination of this approach with omics analyses is a powerful tool for functional genomics. Recently, a global gene expression analysis of maritime pine seedlings in which the *PpNAC1* transcription factor had been silenced allowed the elucidation of the role of this gene in wood formation and the relationships of this gene with other gene expression regulators involved in this process (Pascual et al., 2018). Notably, gene silencing has been very useful for studying the phenylpropanoid synthesis pathway in the context of lignification in conifers. This method has allowed the functional characterization of each member of the pathway, highlighting the roles of these members and quantifying their importance (Möller, Steward, Phillips, Flint, & Wagner, 2005; Trontin et al., 2014; Wadenbäck et al., 2008; Wagner et al., 2009; Wagner, Phillips, Narayan, Moody, & Geddes, 2005; Wagner et al., 2007, 2013, 2011).

Additionally, different studies have used genetic transformation to over-express genes (e.g. Dalman et al., 2017; Mageroy et al., 2017; Möller, McDonald, Walter, & Harris, 2003; Nagel, Berasategui, Paetz, Gershenzon, & Schmidt, 2014; Uddenberg, Abrahamsson, & von Arnold, 2016) and to study the functions of gene promoters (Bedon, Levasseur, Grima-Pettenati, Seguin, & MacKay, 2009). Furthermore, to study gene function, genetic manipulation can also provide a biotechnological opportunity to generate better tree lines with improved characteristics. For example, the transformation of tracheary elements of *P. taeda* with genes encoding ferulate 5-hydroxylase (F5H) and caffeic acid O-methyltransferase (COMT) from *Liquidambar styraciflua* resulted in the production of lignin containing sinapyl alcohol, which is a type of monolignol not found in conifers making this wood more appropriated for biofuel production (Wagner et al., 2015).

Despite the great potential of these techniques, their use to study conifers is still scarce due to their unfavourable characteristics, mainly their

long-life cycles. As a resource, to avoid or minimize this challenge, researchers have used cultured cell lines of conifers to perform genetic manipulations and analyse these cultures (*e.g.* Möller et al., 2005), developing somatic embryos (*e.g.* Klimaszewska, Hargreaves, Lelu-Walter, & Trontin, 2016; Pascual et al., 2018), and *in vitro* tissues as tracheary elements (*e.g.* Wagner et al., 2005, 2015). Additionally, transient gene expression in protoplasts has been employed as a successful alternative to study conifer gene promoters (Craven-Bartle et al., 2013; Gómez-Maldonado, Cánovas, & Ávila, 2004; Gómez-Maldonado, Crespillo, Ávila, & Cánovas, 2001; Pascual et al., 2018; Rueda-López, Crespillo, Cánovas, & Ávila, 2008). A widely used strategy to analyse conifer genes *via* genetic manipulation is overexpression in heterologous systems such as *Arabidopsis*, tobacco or poplar (*e.g.* Gao et al., 2013; Hills, Khan, & López-Juez, 2015; Katahata, Futamura, Igasaki, & Shinohara, 2014; Rueda-López, Cañas, Canales, Cánovas, & Ávila, 2015; Rueda-López et al., 2017; Rupps et al., 2016). Complementation of *Arabidopsis* mutants with the putative homologue of the conifer gene under study is a commonly used approach (*e.g.* Karlgren, Gyllenstrand, Källman, & Lagercrantz, 2013; Larsson, Sundström, Sitbon, & Von Arnold, 2012; Li et al., 2013; Maloney et al., 2012; Nilsson, Carlsbecker, Sundås-Larsson, & Vahala, 2007). Alternatively, complementation of *E. coli* (*e.g.* Hosoo, Kimura, Nanatani, & Uozumi, 2014) and yeast mutants (*e.g.* Castro-Rodríguez et al., 2016) when possible is extremely useful. Mutant complementation in heterologous systems provides functional information about the genes and the products of these genes (*e.g.* El-Azaz, de la Torre, Ávila, & Cánovas, 2016), while overexpression is more adequate as a biotechnological tool (*e.g.* Castro-Rodríguez et al., 2016).

One way to overcome the problems associated with the generation of mutant and transgenic conifers would be the use of genome editing techniques such as CRISPR/Cas9 technology (Charpentier & Doudna, 2013; Doudna & Charpentier, 2014; Feng et al., 2013). This method can potentially be used for directed mutation, modification or substitution of specific genes, facilitating genetic studies of gene function modification (loss, reduction, improvement or shift, Fig. 5) (Nonaka, Arai, Takayama, Matsukura, & Ezura, 2017; Wang et al., 2014) and differential gene expression (Park, Dempewolf, Zhang, & Wang, 2017; Shi et al., 2017). The opportunities are diverse, and this method has the advantage of being a targeting method. Unfortunately, despite the promise of this method, there are no published studies using CRISPR/Cas9 technology in conifers. Thus, CRISPR/Cas9 seems to hold great potential for conifer research.

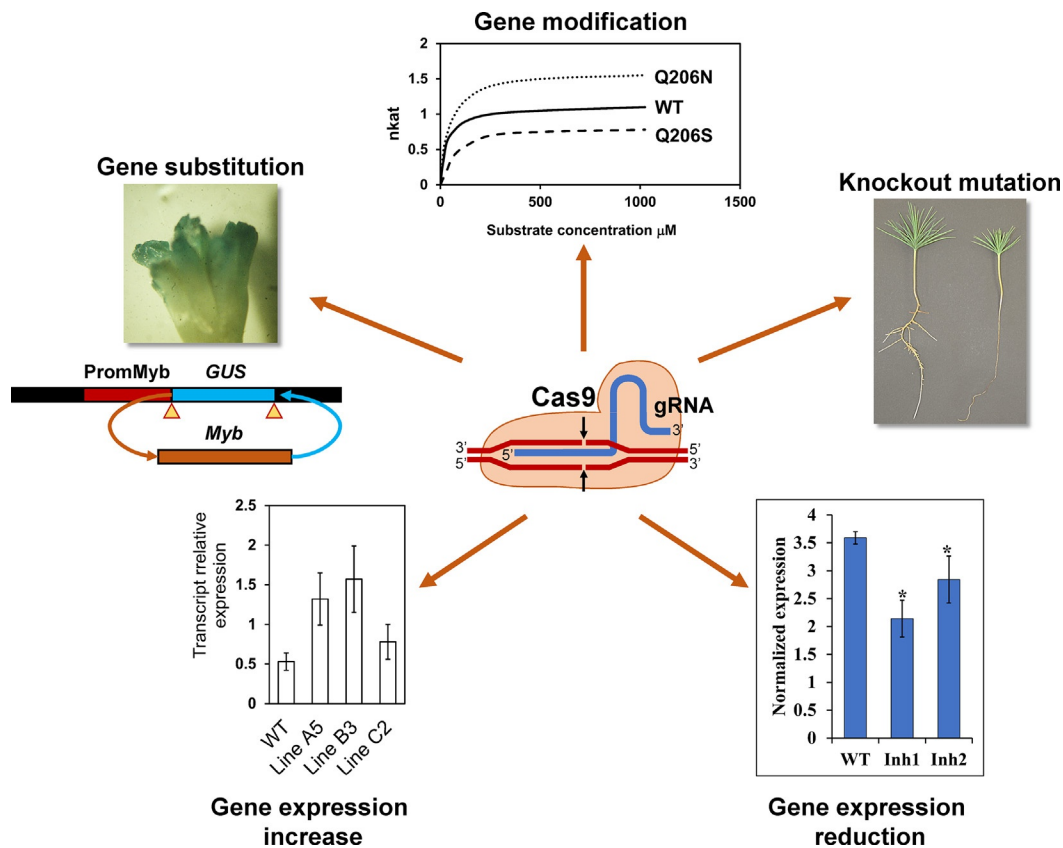


Fig. 5 Potential uses of gene editing methods (as CRISPR/CAS9 system) in conifers.



3. Conclusions and perspectives

Overall, functional genomic research in conifers has advanced enormously over the last 15–20 years following the development of omics techniques. Considering the main obstacle, which is the unfavourable characteristics of conifers as experimental models, the accumulation of data and the generation of scientific resources for conifer functional genomics are impressive achievements. Among these achievements, the assembly of the whole-genome sequences of several conifer species is the most relevant. Maybe, in the future, this achievement will be crucial for obtaining further information of biological and biotechnological value, although in order to reach that goal, improvement of the assemblies and annotations is mandatory. Current, and maybe future, techniques to study the regulation of gene expression and chromatin dynamics require high-quality whole-genome sequences.

At the genetic level, functional studies have been limited since the mutant screening and transgenesis are difficult to achieve. However, the application of genome editing techniques opens new perspectives in the field. When the possibility of manipulating the functions of conifer genes in a “surgical” way is realized, such studies will increase and profoundly change our knowledge about conifer biology. All these new resources are the “holy grail” towards obtaining trees with improved characteristics.

Acknowledgements

R.A.C. was supported by a grant from the Spanish *Ministerio de Economía y Competitividad* (BIO2015-73512-JIN; MINECO/AEI/FEDER, UE). Research work in the authors’s laboratory was supported by grants from *Ministerio de Economía y Competitividad* (BIO2015-69285-R) and the European Union (FP7-KBBE-289841).

References

- Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., et al. (1991). Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science*, 252, 1651–1656. <https://doi.org/10.1126/science.2047873>.
- Allona, I., Quinn, M., Shoop, E., Swope, K., St Cyr, S., Carlis, J., et al. (1998). Analysis of xylem formation in pine by cDNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 9693–9698. <https://doi.org/10.1073/pnas.95.16.9693>.
- Alonso, P., Cortizo, M., Cantón, F. R., Fernández, B., Rodríguez, A., Centeno, M. L., et al. (2007). Identification of genes differentially expressed during adventitious shoot induction in *Pinus pinea* cotyledons by subtractive hybridization and quantitative PCR. *Tree Physiology*, 27, 1721–1730. <https://doi.org/10.1093/treephys/27.12.1721>.

- Alonso, J. M., & Stepanova, A. N. (2015). Plant functional genomics. In J. M. Alonso & A. N. Stepanova (Eds.), *Vol. 1284. Plant functional genomics: Methods and protocols* (2nd ed.). New York NY: Springer New York. <https://doi.org/10.1007/978-1-4939-2444-8>.
- Alvarez, J. M., Bueno, N., Cañas, R. A., Avila, C., Cánovas, F. M., & Ordás, R. J. (2018). Analysis of the WUSCHEL-RELATED HOMEODOMAIN gene family in *Pinus pinaster*: New insights into the gene family evolution. *Plant Physiology and Biochemistry*, 123, 304–318. <https://doi.org/10.1016/j.plaphy.2017.12.031>.
- Alvarez, J. M., Sohlberg, J., Engström, P., Zhu, T., Englund, M., Moschou, P. N., et al. (2015). The WUSCHEL-RELATED HOMEODOMAIN 3 gene PaWOX3 regulates lateral organ formation in Norway spruce. *New Phytologist*, 208, 1078–1088. <https://doi.org/10.1111/nph.13536>.
- Aronen, T., & Ryyänänen, L. (2012). Variation in telomeric repeats of Scots pine (*Pinus sylvestris* L.). *Tree Genetics & Genomes*, 8, 267–275. <https://doi.org/10.1007/s11295-011-0438-7>.
- Asiegbu, F. O., Nahalkova, J., & Li, G. (2005). Pathogen-inducible cDNAs from the interaction of the root rot fungus *Heterobasidion annosum* with Scots pine (*Pinus sylvestris* L.). *Plant Science*, 168, 365–372. <https://doi.org/10.1016/j.plantsci.2004.08.010>.
- Badenes, M. L., Fernández i Martí, A., Ríos, G., & Rubio-Cabetas, M. J. (2016). Application of genomic technologies to the breeding of trees. *Frontiers in Genetics*, 7, 198. <https://doi.org/10.3389/fgene.2016.00198>.
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Research*, 41, 991–995. <https://doi.org/10.1093/nar/gks1193>.
- Bedon, F., Bomal, C., Caron, S., Levasseur, C., Boyle, B., Mansfield, S. D., et al. (2010). Subgroup 4 R2R3-MYBs in conifer trees: Gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses. *Journal of Experimental Botany*, 61, 3847–3864. <https://doi.org/10.1093/jxb/erq196>.
- Bedon, F., Grima-Pettenati, J., & Mackay, J. (2007). Conifer R2R3-MYB transcription factors: Sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). *BMC Plant Biology*, 7, 17. <https://doi.org/10.1186/1471-2229-7-17>.
- Bedon, F., Levasseur, C., Grima-Pettenati, J., Seguin, A., & MacKay, J. (2009). Sequence analysis and functional characterization of the promoter of the *Picea glauca* cinnamyl alcohol dehydrogenase gene in transgenic white spruce plants. *Plant Cell Reports*, 28, 787–800. <https://doi.org/10.1007/s00299-009-0688-0>.
- Benson, G. (2017). Editorial: The 15th annual nucleic acids research web server issue 2017. *Nucleic Acids Research*, 45, W1–W5. <https://doi.org/10.1093/nar/gkx457>.
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2017). GenBank. *Nucleic Acids Research*, 45, D37–D42. <https://doi.org/10.1093/nar/gkw1070>.
- Bérubé, Y., Zhuang, J., Rungis, D., Ralph, S., Bohlmann, J., & Ritland, K. (2007). Characterization of EST-SSRs in loblolly pine and spruce. *Tree Genetics & Genomes*, 3, 251–259. <https://doi.org/10.1007/s11295-006-0061-1>.
- Birol, I., Raymond, A., Jackman, S. D., Pleasance, S., Coope, R., Taylor, G. A., et al. (2013). Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics*, 29, 1492–1497. <https://doi.org/10.1093/bioinformatics/btt178>.
- Boch, J. (2011). TALEs of genome targeting. *Nature Biotechnology*, 29, 135–136. <https://doi.org/10.1038/nbt.1767>.
- Bomal, C., Bedon, F., Caron, S., Mansfield, S. D., Levasseur, C., Cooke, J. E. K., et al. (2008). Involvement of *Pinus taeda* MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: A comparative in planta analysis. *Journal of Experimental Botany*, 59, 3925–3939. <https://doi.org/10.1093/jxb/ern234>.

- Bomal, C., Duval, I., Giguère, I., Fortin, E., Caron, S., Stewart, D., et al. (2014). Opposite action of R2R3-MYBs from different subgroups on key genes of the shikimate and monolignol pathways in spruce. *Journal of Experimental Botany*, 65, 495–508. <https://doi.org/10.1093/jxb/ert398>.
- Bravo, S., Bertin, A., Turner, A., Sepúlveda, F., Jopia, P., Parra, M. J., et al. (2017). Differences in DNA methylation, DNA structure and embryogenesis-related gene expression between embryogenic and non embryogenic lines of *Pinus radiata* D. don. *Plant Cell, Tissue and Organ Culture*, 130, 521–529. <https://doi.org/10.1007/s11240-017-1242-3>.
- Cairney, J., Zheng, L., Cowels, A., Hsiao, J., Zismann, V., Liu, J., et al. (2006). Expressed sequence tags from loblolly pine embryos reveal similarities with angiosperm embryogenesis. *Plant Molecular Biology*, 62, 485–501. <https://doi.org/10.1007/s11103-006-9035-9>.
- Canales, J., Bautista, R., Label, P., Gómez-Maldonado, J., Lesur, I., Fernández-Pozo, N., et al. (2014). *De novo* assembly of maritime pine transcriptome: Implications for forest breeding and biotechnology. *Plant Biotechnology Journal*, 12, 286–299. <https://doi.org/10.1111/pbi.12136>.
- Cañas, R. A., Canales, J., Muñoz-Hernández, C., Granados, J. M., Ávila, C., García-Martín, M. L., et al. (2015). Understanding developmental and adaptive cues in pine through metabolite profiling and co-expression network analysis. *Journal of Experimental Botany*, 66, 3113–3127. <https://doi.org/10.1093/jxb/erv118>.
- Cañas, R. A., de la Torre, F., Pascual, M., Avila, C., & Cánovas, F. (2016). Nitrogen economy and nitrogen environmental interactions in conifers. *Agronomy*, 6, 26. <https://doi.org/10.3390/agronomy6020026>.
- Cañas, R. A., Feito, I., Fuente-Maqueda, J., Avila, C., Majada, J., & Canovas, F. (2015). Transcriptome-wide analysis supports environmental adaptations of two *Pinus pinaster* populations from contrasting habitats. *BMC Genomics*, 16, 909. <https://doi.org/10.1186/s12864-015-2177-x>.
- Cañas, R. A., Pascual, M. B., Castro-Rodríguez, V., Ávila, C., Sterck, L., Van de Peer, Y., et al. (2017). The gene expression landscape of pine seedling tissues. *The Plant Journal*, 91, 1064–1087. <https://doi.org/10.1111/tpj.13617>.
- Carlsbecker, A., Sundström, J. F., Englund, M., Uddenberg, D., Izquierdo, L., Kvarnheden, A., et al. (2013). Molecular control of normal and *acrocona* mutant seed cone development in Norway spruce (*Picea abies*) and the evolution of conifer ovule-bearing organs. *New Phytologist*, 200, 261–275. <https://doi.org/10.1111/nph.12360>.
- Castro-Rodríguez, V., Assaf-Casals, I., Pérez-Tienda, J., Fan, X., Avila, C., Miller, A., et al. (2016). Deciphering the molecular basis of ammonium uptake and transport in maritime pine. *Plant, Cell & Environment*, 39, 1669–1682. <https://doi.org/10.1111/pce.12692>.
- Castro-Rodríguez, V., García-Gutiérrez, A., Canales, J., Cañas, R. A., Kirby, E. G., Avila, C., et al. (2016). Poplar trees for phytoremediation of high levels of nitrate and applications in bioenergy. *Plant Biotechnology Journal*, 14, 299–312. <https://doi.org/10.1111/pbi.12384>.
- Chancerel, E., Lamy, J. B., Lesur, I., Noirot, C., Klopp, C., Ehrenmann, F., et al. (2013). High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. *BMC Biology*, 11, 50. <https://doi.org/10.1186/1741-7007-11-50>.
- Chancerel, E., Lepoittevin, C., Le Provost, G., Lin, Y., Jaramillo-correa, J. P., Eckert, A. J., et al. (2011). Development and implementation of a highly multiplexed SNP array for genetic mapping in maritime pine and comparative mapping with loblolly pine. *BMC Genomics*, 12, 368. <https://doi.org/10.1186/1471-2164-12-368>.
- Charpentier, E., & Doudna, J. A. (2013). Biotechnology: Rewriting a genome. *Nature*, 495, 50–51. <https://doi.org/10.1038/495050a>.

- Craven-Bartle, B., Pascual, M. B., Cánovas, F. M., & Ávila, C. (2013). A Myb transcription factor regulates genes of the phenylalanine pathway in maritime pine. *Plant Journal*, 74, 755–766. <https://doi.org/10.1111/tpj.12158>.
- Dalman, K., Wind, J. J., Nemesio-Gorritz, M., Hammerbacher, A., Lundén, K., Ezcurra, I., et al. (2017). Overexpression of *PaNAC03*, a stress induced NAC gene family transcription factor in Norway spruce leads to reduced flavonol biosynthesis and aberrant embryo development. *BMC Plant Biology*, 17, 6. <https://doi.org/10.1186/s12870-016-0952-8>.
- De La Torre, A. R., Birol, I., Bousquet, J., Ingvarsson, P. K., Jansson, S., Jones, S. J. M., et al. (2014). Insights into conifer giga-genomes. *Plant Physiology*, 166, 1724–1732. <https://doi.org/10.1104/pp.114.248708>.
- de Vega-Bartol, J. J., Simões, M., Lorenz, W., Rodrigues, A. S., Alba, R., Dean, J. F. D., et al. (2013). Transcriptomic analysis highlights epigenetic and transcriptional regulation during zygotic embryo development of *Pinus pinaster*. *BMC Plant Biology*, 13, 123. <https://doi.org/10.1186/1471-2229-13-123>.
- Domec, J. C., Lachenbruch, B., Meinzer, F. C., Woodruff, D. R., Warren, J. M., & McCulloh, K. A. (2008). Maximum height in a conifer is associated with conflicting requirements for xylem design. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 12069–12074. <https://doi.org/10.1073/pnas.0710418105>.
- Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR–Cas9. *Science*, 346, 1258096. <https://doi.org/10.1126/science.1258096>.
- Dufva, M. (2009). Introduction to microarray technology. In M. Dufva (Ed.), *DNA microarrays for biomedical research: Methods and protocols* (pp. 1–22). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-59745-538-1_1.
- El-Azaz, J., de la Torre, F., Ávila, C., & Cánovas, F. M. (2016). Identification of a small protein domain present in all plant lineages that confers high prephenate dehydratase activity. *Plant Journal*, 87, 215–229. <https://doi.org/10.1111/tpj.13195>.
- Essen, P. A., Ehnström, B., Ericsson, L., & Sjöberg, K. (1997). Boreal forest. *Ecological Bulletins*, 46, 16–47. <http://www.jstor.org/stable/20113207>.
- Farjon, A. (2010). *A handbook of the world's conifers*, Vol. 2. Leiden–Boston: Brill. <https://doi.org/10.1163/9789047430629>.
- Farjon, A., & Filer, D. (2013). *An atlas of the world's conifers: An analysis of their distribution, biogeography, diversity and conservation status*. Leiden–Boston: Brill. <https://doi.org/10.1017/CBO9781107415324.004>.
- Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D. L., Wei, P., et al. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research*, 23, 1229–1232. <https://doi.org/10.1038/cr.2013.114>.
- Fernández-Pozo, N., Canales, J., Guerrero-Fernández, D., Villalobos, D. P., Díaz-Moreno, S. M., Bautista, R., et al. (2011). EuroPineDB: A high-coverage web database for maritime pine transcriptome. *BMC Genomics*, 12, 366. <https://doi.org/10.1186/1471-2164-12-366>.
- Flanary, B. E., & Kletetschka, G. (2005). Analysis of telomere length and telomerase activity in tree species of various life-spans, and with age in the bristlecone pine *Pinus longaeva*. *Biogerontology*, 6, 101–111. <https://doi.org/10.1007/s10522-005-3484-4>.
- Fossdal, C. G., Yaqoob, N., Krokene, P., Kvaalen, H., Solheim, H., & Yakovlev, I. A. (2012). Local and systemic changes in expression of resistance genes, *nb-lrr* genes and their putative microRNAs in Norway spruce after wounding and inoculation with the pathogen *Ceratocystis polonica*. *BMC Plant Biology*, 12, 105. <https://doi.org/10.1186/1471-2229-12-105>.
- Futamura, N., Totoki, Y., Toyoda, A., Igasaki, T., Nanjo, T., Seki, M., et al. (2008). Characterization of expressed sequence tags from a full-length enriched cDNA library of *Cryptomeria japonica* male strobili. *BMC Genomics*, 9, 383. <https://doi.org/10.1186/1471-2164-9-383>.

- Galperin, M. Y., Fernández-Suárez, X. M., & Rigden, D. J. (2017). The 24th annual nucleic acids research database issue: A look back and upcoming changes. *Nucleic Acids Research*, 45, D1–D11. <https://doi.org/10.1093/nar/gkw1188>.
- Gao, Y., Bian, L., Shi, J., Xu, J., Xi, M., & Wang, G. (2013). Expression of a conifer *COBRA*-like gene *CICOBL1* from Chinese fir (*Cunninghamia lanceolata*) alters the leaf architecture in tobacco. *Plant Physiology and Biochemistry*, 70, 483–491. <https://doi.org/10.1016/j.plaphy.2013.06.013>.
- García-Gutiérrez, A., Cánovas, F. M., & Avila, C. (2018). Glutamate synthases from conifers: Gene structure and phylogenetic studies. *BMC Genomics*, 19, 65. <https://doi.org/10.1186/s12864-018-4454-y>.
- Gibson, R., Alako, B., Amid, C., Cerdeño-Tárraga, A., Cleland, I., Goodgame, N., et al. (2016). Biocuration of functional annotation at the European nucleotide archive. *Nucleic Acids Research*, 44, D58–D66. <https://doi.org/10.1093/nar/gkv1311>.
- Gómez-Maldonado, J., Cánovas, F. M., & Avila, C. (2004). Molecular analysis of the 5'-upstream region of a gibberellin-inducible cytosolic glutamine synthetase gene (GS1b) expressed in pine vascular tissue. *Planta*, 218, 1036–1045. <https://doi.org/10.1007/s00425-003-1185-2>.
- Gómez-Maldonado, J., Crespillo, R., Ávila, C., & Cánovas, F. M. (2001). Efficient preparation of maritime pine (*Pinus pinaster*) protoplasts suitable for transgene expression analysis. *Plant Molecular Biology Reporter*, 19, 361–366. <https://doi.org/10.1007/BF02772834>.
- Guo, A., Zheng, C. X., & Yang, Y. Y. (2014). Differential expression of *SLOW WALKER2* homologue in ovules of female sterile mutant and fertile clone of *Pinus tabulaeformis*. *Russian Journal of Developmental Biology*, 45, 78–84. <https://doi.org/10.1134/S1062360414020052>.
- Han, Q., Shinohara, K., Kakubari, Y., & Mukai, Y. (2003). Photoprotective role of rhodoxanthin during cold acclimation in *Cryptomeria japonica*. *Plant, Cell and Environment*, 26, 715–723. <https://doi.org/10.1046/j.1365-3040.2003.01008.x>.
- Harfouche, A., Meilan, R., Kirst, M., Morgante, M., Boerjan, W., Sabatti, M., et al. (2012). Accelerating the domestication of forest trees in a changing world. *Trends in Plant Science*, 17, 64–72. <https://doi.org/10.1016/j.tplants.2011.11.005>.
- Hedman, H., Zhu, T., von Arnold, S., & Sohlberg, J. J. (2013). Analysis of the WUSCHEL-RELATED HOMEBOX gene family in the conifer *Picea abies* reveals extensive conservation as well as dynamic patterns. *BMC Plant Biology*, 13, 89. <https://doi.org/10.1186/1471-2229-13-89>.
- Hills, A. C., Khan, S., & López-Juez, E. (2015). Chloroplast biogenesis-associated nuclear genes: Control by plastid signals evolved prior to their regulation as part of photomorphogenesis. *Frontiers in Plant Science*, 6, 1078. <https://doi.org/10.3389/fpls.2015.01078>.
- Hirao, T., Watanabe, A., Kurita, M., Kondo, T., & Takata, K. (2009). A frameshift mutation of the chloroplast *matK* coding region is associated with chlorophyll deficiency in the *Cryptomeria japonica* virescent mutant *Wogon-Sugi*. *Current Genetics*, 55, 311–321. <https://doi.org/10.1007/s00294-009-0247-9>.
- Hosoo, Y., Kimura, Y., Nanatani, K., & Uozumi, N. (2014). Molecular cloning and expression analysis of a gene encoding KUP/HAK/KT-type potassium uptake transporter from *Cryptomeria japonica*. *Trees*, 28, 1527–1537. <https://doi.org/10.1007/s00468-014-1059-1>.
- Hu, X. G., Liu, H., Jin, Y. Q., Sun, Y. Q., Li, Y., Zhao, W., et al. (2016). *De novo* transcriptome assembly and characterization for the widespread and stress-tolerant conifer *Platycladus orientalis*. *PLoS One*, 11, e0148985. <https://doi.org/10.1371/journal.pone.0148985>.
- Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 316, 1497–1502. <https://doi.org/10.1126/science.1141319>.

- Jokipii-Lukkari, S., Sundell, D., Nilsson, O., Hvidsten, T. R., Street, N. R., & Tuominen, H. (2017). NorWood: A gene expression resource for evo-devo studies of conifer wood development. *New Phytologist*, 216, 482–494. <https://doi.org/10.1111/nph.14458>.
- Karlgrén, A., Gyllenstrand, N., Källman, T., & Lagercrantz, U. (2013). Conserved function of core clock proteins in the gymnosperm Norway spruce (*Picea abies* L. Karst). *PLoS One*, 8, e60110. <https://doi.org/10.1371/journal.pone.0060110>.
- Katahata, S. I., Futamura, N., Igasaki, T., & Shinohara, K. (2014). Functional analysis of SOC1-like and AGL6-like MADS-box genes of the gymnosperm *Cryptomeria japonica*. *Tree Genetics & Genomes*, 10, 317–327. <https://doi.org/10.1007/s11295-013-0686-9>.
- Kirst, M., Johnson, A. F., Baucom, C., Ulrich, E., Hubbard, K., Staggs, R., et al. (2003). Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 7383–7388. <https://doi.org/10.1073/pnas.1132171100>.
- Kjellsen, T. D., Shiryayeva, L., Schröder, W. P., & Strimbeck, G. R. (2010). Proteomics of extreme freezing tolerance in Siberian spruce (*Picea obovata*). *Journal of Proteomics*, 73, 965–975. <https://doi.org/10.1016/j.jpro.2009.12.010>.
- Klimaszewska, K., Hargreaves, C., Lelu-Walter, M. A., & Trontin, J. F. (2016). Advances in conifer somatic embryogenesis since year 2000. *Methods in Molecular Biology*, 1359, 131–166. https://doi.org/10.1007/978-1-4939-3061-6_7.
- Klimaszewska, K., Lachance, D., Pelletier, G., Lelu, M. A., & Seguin, A. (2001). Regeneration of transgenic *Picea Glauca*, *P. Mariana*, and *P. Abies* after cocultivation of embryogenic tissue with *Agrobacterium tumefaciens*. *In Vitro Cellular & Developmental Biology—Plant*, 37, 748–755. <https://doi.org/10.1079/IVP2001230>.
- Lara, A., & Villalba, R. (1993). A 3620-year temperature record from *Fitzroya cupressoides* tree rings in southern South America. *Science*, 260, 1104–1106. <https://doi.org/10.1126/science.260.5111.1104>.
- Larsson, E., Sundström, J. F., Sitbon, F., & Von Arnold, S. (2012). Expression of *PaNAC01*, a *Picea abies* CUP-SHAPED COTYLEDON orthologue, is regulated by polar auxin transport and associated with differentiation of the shoot apical meristem and formation of separated cotyledons. *Annals of Botany*, 110, 923–934. <https://doi.org/10.1093/aob/mcS151>.
- Lee, C. H., Chan, M. H., Wang, Y. N., & Chu, F. H. (2006). Gene investigation into the inner bark of Taiwan (*Taiwania cryptomerioides*). *Botanical Studies*, 47, 111–118.
- Li, Z., De La Torre, A., Sterck, L., Cánovas, F. M., Avila, C., Merino, I., et al. (2017). Single-copy genes as molecular markers for phylogenomic studies in seed plants. *Genome Biology and Evolution*, 9, 1130–1147. <https://doi.org/10.1093/gbe/evx070>.
- Li, P. C. H., Sedighi, A., & Wang, L. (2016). In P. C. H. Li, A. Sedighi, & L. Wang (Eds.), Vol. 1368. *Microarray Technology*. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4939-3136-1>.
- Li, X., Wu, H. X., Dillon, S. K., & Southerton, S. G. (2009). Generation and analysis of expressed sequence tags from six developing xylem libraries in *Pinus radiata* D. Don. *BMC Genomics*, 10, 41. <https://doi.org/10.1186/1471-2164-10-41>.
- Li, L., Yu, Y., Wei, J., Huang, G., Zhang, D., Liu, Y., et al. (2013). Homologous HAP5 subunit from *Picea wilsonii* improved tolerance to salt and decreased sensitivity to ABA in transformed *Arabidopsis*. *Planta*, 238, 345–356. <https://doi.org/10.1007/s00425-013-1894-0>.
- Lippert, D., Yuen, M., & Bohlmann, J. (2009). Spruce proteome DB: A resource for conifer proteomics research. *Tree Genetics and Genomes*, 5, 723–727. <https://doi.org/10.1007/s11295-009-0220-2>.

- Liu, Y., & El-Kassaby, Y. A. (2017). Landscape of fluid sets of hairpin-derived 21-/24-nt-long small RNAs at seed set uncovers special epigenetic features in *Picea glauca*. *Genome Biology and Evolution*, 9, 82–92. <https://doi.org/10.1093/gbe/evw283>.
- Liu, J. J., Schoettle, A. W., Snieszko, R. A., Sturrock, R. N., Zamany, A., Williams, H., et al. (2016). Genetic mapping of *Pinus flexilis* major gene (Cr4) for resistance to white pine blister rust using transcriptome-based SNP genotyping. *BMC Genomics*, 17, 753. <https://doi.org/10.1186/s12864-016-3079-2>.
- Lorenz, W. W., Ayyampalayam, S., Bordeaux, J. M., Howe, G. T., Jermstad, K. D., Neale, D. B., et al. (2012). Conifer DBMagic: A database housing multiple *de novo* transcriptome assemblies for 12 diverse conifer species. *Tree Genetics & Genomes*, 8, 1477–1485. <https://doi.org/10.1007/s11295-012-0547-y>.
- Lorenz, K., & Lal, R. (2010). *Carbon sequestration in forest ecosystems*. Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-90-481-3266-9>.
- Lorenz, W. W., Sun, F., Liang, C., Kolychev, D., Wang, H., Zhao, X., et al. (2006). Water stress-responsive genes in loblolly pine (*Pinus taeda*) roots identified by analyses of expressed sequence tag libraries. *Tree Physiology*, 26, 1–16. <https://doi.org/10.1093/treephys/26.1.1>.
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS Computational Biology*, 13, e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>.
- Lu, M., Krutovsky, K. V., Nelson, C. D., Koralewski, T. E., Byram, T. D., & Loopstra, C. A. (2016). Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.). *BMC Genomics*, 17, 730. <https://doi.org/10.1186/s12864-016-3081-8>.
- Lu, M., Krutovsky, K. V., Nelson, C. D., West, J. B., Reilly, N. A., & Loopstra, C. A. (2017). Association genetics of growth and adaptive traits in loblolly pine (*Pinus taeda* L.) using whole-exome-discovered polymorphisms. *Tree Genetics & Genomes*, 13, 57. <https://doi.org/10.1007/s11295-017-1140-1>.
- MacKay, J., Dean, J. F. D., Plomion, C., Peterson, D. G., Cánovas, F. M., Pavy, N., et al. (2012). Towards decoding the conifer giga-genome. *Plant Molecular Biology*, 80, 555–569. <https://doi.org/10.1007/s11103-012-9961-7>.
- MacKay, J. J., O'Malley, D. M., Presnell, T., Booker, F. L., Campbell, M. M., Whetten, R. W., et al. (1997). Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 8255–8260.
- Mageroy, M. H., Lachance, D., Jancsik, S., Parent, G., Séguin, A., Mackay, J., et al. (2017). *In vivo* function of *Pgβglu-1* in the release of acetophenones in white spruce. *PeerJ*, 5, e3535. <https://doi.org/10.7717/peerj.3535>.
- Maloney, V. J., Samuels, A. L., & Mansfield, S. D. (2012). The endo-1,4-β-glucanase *Korrigan* exhibits functional conservation between gymnosperms and angiosperms and is required for proper cell wall formation in gymnosperms. *New Phytologist*, 193, 1076–1087. <https://doi.org/10.1111/j.1469-8137.2011.03998.x>.
- Mann, I. K., Wegrzyn, J. L., & Rajora, O. P. (2013). Generation, functional annotation and comparative analysis of black spruce (*Picea mariana*) ESTs: An important conifer genomic resource. *BMC Genomics*, 14, 702. <https://doi.org/10.1186/1471-2164-14-702>.
- Martin, S. A. M., Dehler, C. E., & Król, E. (2016). Transcriptomic responses in the fish intestine. *Developmental and Comparative Immunology*, 64, 103–117. <https://doi.org/10.1016/j.dci.2016.03.014>.
- Mashima, J., Kodama, Y., Kosuge, T., Fujisawa, T., Katayama, T., Nagasaki, H., et al. (2016). DNA data bank of Japan (DDBJ) progress report. *Nucleic Acids Research*, 44, D51–D57. <https://doi.org/10.1093/nar/gkv1105>.

- Meyer, C. A., & Liu, X. S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nature Reviews. Genetics*, 15, 709–721. <https://doi.org/10.1038/nrg3788>.
- Möller, R., McDonald, A. G., Walter, C., & Harris, P. J. (2003). Cell differentiation, secondary cell-wall formation and transformation of callus tissue of *Pinus radiata* D. Don. *Planta*, 217, 736–747. <https://doi.org/10.1007/s00425-003-1053-0>.
- Möller, R., Steward, D., Phillips, L., Flint, H., & Wagner, A. (2005). Gene silencing of cinnamyl alcohol dehydrogenase in *Pinus radiata* callus cultures. *Plant Physiology and Biochemistry*, 43, 1061–1066. <https://doi.org/10.1016/j.plaphy.2005.11.001>.
- Morin, R. D., Aksay, G., Dolgosheina, E., Ebhardt, H. A., Magrini, V., Mardis, E. R., et al. (2008). Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. *Genome Research*, 18, 571–584. <https://doi.org/10.1101/gr.6897308>.
- Moschou, P. N., Savenkov, E. I., Minina, E. A., Fukada, K., Reza, S. H., Gutierrez-Beltran, E., et al. (2016). *EXTRA SPINDLE POLES* (Separase) controls anisotropic cell expansion in Norway spruce (*Picea abies*) embryos independently of its role in anaphase progression. *New Phytologist*, 212, 232–243. <https://doi.org/10.1111/nph.14012>.
- Nagel, R., Berasategui, A., Paetz, C., Gershenzon, J., & Schmidt, A. (2014). Overexpression of an isoprenyl diphosphate synthase in spruce leads to unexpected terpene diversion products that function in plant defense. *Plant Physiology*, 164, 555–569. <https://doi.org/10.1104/pp.113.228940>.
- Naithani, S., Preece, J., D'Eustachio, P., Gupta, P., Amarasinghe, V., Dharmawardhana, P. D., et al. (2017). Plant reactome: A resource for plant pathways and comparative analysis. *Nucleic Acids Research*, 45, D1029–D1039. <https://doi.org/10.1093/nar/gkw932>.
- Neale, D. B., McGuire, P. E., Wheeler, N. C., Stevens, K. A., Crepeau, M. W., Cardeno, C., et al. (2017). The Douglas-fir genome sequence reveals specialization of the photosynthetic apparatus in pinaceae. *G3 (Bethesda)*, 7, 3157–3167. <https://doi.org/10.1534/g3.117.300078>.
- Neale, D. B., Wegrzyn, J. L., Stevens, K. A., Zimin, A. V., Puiu, D., Crepeau, M. W., et al. (2014). Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology*, 15, R59. <https://doi.org/10.1186/gb-2014-15-3-r59>.
- Netotea, S., Sundell, D., Street, N. R., & Hvidsten, T. R. (2014). ComPIEx: Conservation and divergence of co-expression networks in *A. thaliana*, *Populus* and *O. sativa*. *BMC Genomics*, 15, 106. <https://doi.org/10.1186/1471-2164-15-106>.
- Neves, L. G., Davis, J. M., Barbazuk, W. B., & Kirst, M. (2013). Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal*, 75, 146–156. <https://doi.org/10.1111/tpj.12193>.
- Neves, L. G., Davis, J. M., Barbazuk, W. B., & Kirst, M. (2014). A high-density gene map of loblolly pine (*Pinus taeda* L.) based on exome sequence capture genotyping. *G3 (Bethesda)*, 4, 29–37. <https://doi.org/10.1534/g3.113.008714>.
- Nilsson, L., Carlsbecker, A., Sundås-Larsson, A., & Vahala, T. (2007). *APETALA2* like genes from *Picea abies* show functional similarities to their *Arabidopsis* homologues. *Planta*, 225, 589–602. <https://doi.org/10.1007/s00425-006-0374-1>.
- Ning, D. L., Lu, T. C., Liu, G. F., Yang, C. P., & Wang, B. C. (2013). Proteomic analysis points to a role for RAD23 in apical dominance in *Pinus sylvestris* var. *Mongolica*. *Plant Molecular Biology Reporter*, 31, 1283–1292. <https://doi.org/10.1007/s11105-013-0593-0>.
- Niu, S. H., Liu, C., Yuan, H.-W., Li, P., Li, Y., & Li, W. (2015). Identification and expression profiles of sRNAs and their biogenesis and action-related genes in male and female cones of *Pinus tabuliformis*. *BMC Genomics*, 16, 693. <https://doi.org/10.1186/s12864-015-1885-6>.

- Niu, S., Yuan, H., Sun, X., Porth, I., Li, Y., El-Kassaby, Y. A., et al. (2016). A transcriptomics investigation into pine reproductive organ development. *New Phytologist*, 209, 1278–1289. <https://doi.org/10.1111/nph.13680>.
- Nonaka, S., Arai, C., Takayama, M., Matsukura, C., & Ezura, H. (2017). Efficient increase of γ -aminobutyric acid (GABA) content in tomato fruits by targeted mutagenesis. *Scientific Reports*, 7, 7057. <https://doi.org/10.1038/s41598-017-06400-y>.
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497, 579–584. <https://doi.org/10.1038/nature12211>.
- O'Malley, R. C., Huang, S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., et al. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*, 165, 1280–1292. <https://doi.org/10.1016/j.cell.2016.04.038>.
- Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: Advances, challenges and opportunities. *Nature Reviews. Genetics*, 12, 87–98. <https://doi.org/10.1038/nrg2934>.
- Pâques, L. E. (2013). In L. E. Pâques (Ed.), *Vol. 25. Forest tree breeding in Europe*. Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-94-007-6146-9>.
- Park, J. J., Dempewolf, E., Zhang, W., & Wang, Z. Y. (2017). RNA-guided transcriptional activation via CRISPR/dCas9 mimics overexpression phenotypes in *Arabidopsis*. *PLoS One*, 12, e0179410. <https://doi.org/10.1371/journal.pone.0179410>.
- Pascual, M., Canovas, F., & Avila, C. (2015). The NAC transcription factor family in maritime pine (*Pinus pinaster*): Molecular regulation of two genes involved in stress responses. *BMC Plant Biology*, 15, 254. <https://doi.org/10.1186/s12870-015-0640-0>.
- Pascual, M. B., El-Azaz, J., de la Torre, F. N., Cañas, R. A., Avila, C., & Cánovas, F. M. (2016). Biosynthesis and metabolic fate of phenylalanine in conifers. *Frontiers in Plant Science*, 7, 1030. <https://doi.org/10.3389/fpls.2016.01030>.
- Pascual, M. B., Llebrés, M. T., Craven-Bartle, B., Cañas, R. A., Cánovas, F. M., & Ávila, C. (2018). *PpNAC1*, a main regulator of phenylalanine biosynthesis and utilization in maritime pine. *Plant Biotechnology Journal*, 16, 1094–1104. <https://doi.org/10.1111/pbi.12854>.
- Pavy, N., Gagnon, F., Deschênes, A., Boyle, B., Beaulieu, J., & Bousquet, J. (2016). Development of highly reliable in silico SNP resource and genotyping assay from exome capture and sequencing: An example from black spruce (*Picea mariana*). *Molecular Ecology Resources*, 16, 588–598. <https://doi.org/10.1111/1755-0998.12468>.
- Pavy, N., Gagnon, F., Rigault, P., Blais, S., Deschênes, A., Boyle, B., et al. (2013). Development of high-density SNP genotyping arrays for white spruce (*Picea glauca*) and transferability to subtropical and nordic congeners. *Molecular Ecology Resources*, 13, 324–336. <https://doi.org/10.1111/1755-0998.12062>.
- Pavy, N., Johnson, J. J., Crow, J. A., Paule, C., Kunau, T., MacKay, J., et al. (2007). ForestTreeDB: A database dedicated to the mining of tree transcriptomes. *Nucleic Acids Research*, 35(Suppl. 1), 888–894. <https://doi.org/10.1093/nar/gkl882>.
- Pavy, N., Lamothe, M., Pelgas, B., Gagnon, F., Birol, I., Bohlmann, J., et al. (2017). 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 Journal*, 90, 189–203. <https://doi.org/10.1111/tpj.13478>.
- Pavy, N., Paule, C., Parsons, L., Crow, J. a., Morency, M.-J., Cooke, J., et al. (2005). Generation, annotation, analysis and database integration of 16,500 white spruce EST clusters. *BMC Genomics*, 6, 144. <https://doi.org/10.1186/1471-2164-6-144>.
- Perry, D. J., & Bousquet, J. (1998). Sequence-tagged-site (STS) markers of arbitrary genes: Development, characterization and analysis of linkage in black spruce. *Genetics*, 149, 1089–1098.
- Pevsner, J. (2015). *Bioinformatics and functional genomics. Vol. 1* (3rd ed.). Chichester, UK: John Wiley & Sons Inc. Retrieved from <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1118581784.html#>.

- Plomion, C., Bastien, C., Bogeat-Triboulot, M. B., Bouffier, L., Déjardin, A., Duplessis, S., et al. (2016). Forest tree genomics: 10 Achievements from the past 10 years and future prospects. *Annals of Forest Science*, 73, 77–103. <https://doi.org/10.1007/s13595-015-0488-3>.
- Proost, S., Van Bel, M., Vaneechoutte, D., Van De Peer, Y., Inzé, D., Mueller-Roeber, B., et al. (2015). PLAZA 3.0: An access point for plant comparative genomics. *Nucleic Acids Research*, 43, D974–D981. <https://doi.org/10.1093/nar/gku986>.
- Provart, N. J., Alonso, J., Assmann, S. M., Bergmann, D., Brady, S. M., Brkljacic, J., et al. (2016). 50 Years of *Arabidopsis* research: Highlights and future directions. *New Phytologist*, 209, 921–944. <https://doi.org/10.1111/nph.13687>.
- Prunier, J., Caron, S., & MacKay, J. (2017). 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*, 18, 97. <https://doi.org/10.1186/s12864-016-3458-8>.
- Qiu, Z., Li, X., Zhao, Y., Zhang, M., Wan, Y., Cao, D., et al. (2015). Genome-wide analysis reveals dynamic changes in expression of microRNAs during vascular cambium development in Chinese fir, *Cunninghamia lanceolata*. *Journal of Experimental Botany*, 66, 3041–3054. <https://doi.org/10.1093/jxb/erv103>.
- Quinn, C. R., Iriyama, R., & Fernando, D. D. (2014). Expression patterns of conserved microRNAs in the male gametophyte of loblolly pine (*Pinus taeda*). *Plant Reproduction*, 27, 69–78. <https://doi.org/10.1007/s00497-014-0241-3>.
- Quinn, C. R., Iriyama, R., & Fernando, D. D. (2015). Computational predictions and expression patterns of conserved microRNAs in loblolly pine (*Pinus taeda*). *Tree Genetics & Genomes*, 11, 806. <https://doi.org/10.1007/s11295-014-0806-1>.
- Raherison, E., Rigault, P., Caron, S., Poulin, P. L., Boyle, B., Verta, J. P., et al. (2012). Transcriptome profiling in conifers and the PiceaGenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. *BMC Genomics*, 13, 434. <https://doi.org/10.1186/1471-2164-13-434>.
- Ralph, S. G., Chun, H., Kolosova, N., Cooper, D., Oddy, C., Ritland, C. E., et al. (2008). A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics*, 9, 484. <https://doi.org/10.1186/1471-2164-9-484>.
- Ralph, J., MacKay, J., Hatfield, R. D., O'Malley, D. M., Whetten, R. W., & Sederoff, R. R. (1997). Abnormal lignin in a loblolly pine mutant. *Science*, 277, 235–239. <https://doi.org/10.1126/science.277.5323.235>.
- Ralph, S. G., Yueh, H., Friedmann, M., Aeschliman, D., Zeznik, J. A., Nelson, C. C., et al. (2006). Conifer defence against insects: Microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale. *Plant, Cell and Environment*, 29, 1545–1570. <https://doi.org/10.1111/j.1365-3040.2006.01532.x>.
- Rigault, P., Boyle, B., Lepage, P., Cooke, J. E. K., Bousquet, J., & MacKay, J. J. (2011). A white spruce gene catalog for conifer genome analyses. *Plant Physiology*, 157, 14–28. <https://doi.org/10.1104/pp.111.179663>.
- Ritland, K. (2012). Genomics of a phylum distant from flowering plants: Conifers. *Tree Genetics and Genomes*, 8, 573–582. <https://doi.org/10.1007/s11295-012-0497-4>.
- Romero-Rodríguez, M. C., Pascual, J., Valledor, L., & Jorrín-Novó, J. (2014). Improving the quality of protein identification in non-model species. Characterization of *Quercus ilex* seed and *Pinus radiata* needle proteomes by using SEQUEST and custom databases. *Journal of Proteomics*, 105, 85–91. <https://doi.org/10.1016/j.jprot.2014.01.027>.
- Rueda-López, M., Cañas, R. A., Canales, J., Cánovas, F. M., & Ávila, C. (2015). The over-expression of the pine transcription factor PpDof5 in *Arabidopsis* leads to increased lignin content and affects carbon and nitrogen metabolism. *Physiologia Plantarum*, 155, 369–383. <https://doi.org/10.1111/ppl.12381>.

- Rueda-López, M., Crespillo, R., Cánovas, F. M., & Avila, C. (2008). Differential regulation of two glutamine synthetase genes by a single Dof transcription factor. *Plant Journal*, 56, 73–85. <https://doi.org/10.1111/j.1365-3113X.2008.03573.x>.
- Rueda-López, M., García-Gutiérrez, A., Cánovas, F. M., & Ávila, C. (2013). The family of Dof transcription factors in pine. *Trees—Structure and Function*, 27, 1547–1557. <https://doi.org/10.1007/s00468-013-0903-z>.
- Rueda-López, M., Pascual, M. B., Pallero, M., Henao, L. M., Lasa, B., Jauregui, I., et al. (2017). Overexpression of a pine Dof transcription factor in hybrid poplars: A comparative study in trees growing under controlled and natural conditions. *PLoS One*, 12, e0174748. <https://doi.org/10.1371/journal.pone.0174748>.
- Rungis, D., Bérubé, Y., Zhang, J., Ralph, S., Ritland, C. E., Ellis, B. E., et al. (2004). Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics*, 109, 1283–1294. <https://doi.org/10.1007/s00122-004-1742-5>.
- Rupps, A., Raschke, J., Rümmler, M., Linke, B., & Zoglauer, K. (2016). Identification of putative homologs of *Larix decidua* to BABYBOOM (BBM), LEAFY COTYLEDON1 (LEC1), WUSCHEL-related HOMEBOX2 (WOX2) and SOMATIC EMBRYO-GENESIS RECEPTOR-like KINASE (SERK) during somatic embryogenesis. *Planta*, 243, 473–488. <https://doi.org/10.1007/s00425-015-2409-y>.
- Seoane-Zonjic, P., Cañas, R. A., Bautista, R., Gómez-Maldonado, J., Arrillaga, I., Fernández-Pozo, N., et al. (2016). Establishing gene models from the *Pinus pinaster* genome using gene capture and BAC sequencing. *BMC Genomics*, 17, 148. <https://doi.org/10.1186/s12864-016-2490-z>.
- Shi, J., Gao, H., Wang, H., Lafitte, H. R., Archibald, R. L., Yang, M., et al. (2017). ARGOS8 variants generated by CRISPR–Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnology Journal*, 15, 207–216. <https://doi.org/10.1111/pbi.12603>.
- Steijger, T., Abril, J. F., Engström, P. G., Kokocinski, F., Abril, J. F., Akerman, M., et al. (2013). Assessment of transcript reconstruction methods for RNA-seq. *Nature Methods*, 10, 1177–1184. <https://doi.org/10.1038/nmeth.2714>.
- Stevens, K. A., Wegrzyn, J. L., Zimin, A., Puiu, D., Crepeau, M., Cardeno, C., et al. (2016). Sequence of the sugar pine megagenome. *Genetics*, 204, 1613–1626. <https://doi.org/10.1534/genetics.116.193227>.
- Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y. C., Sjödin, A., et al. (2015). The plant genome integrative explorer resource: PlantGenIE.org. *New Phytologist*, 208, 1149–1156. <https://doi.org/10.1111/nph.13557>.
- Suren, H., Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Smets, P., Rieseberg, L. H., et al. (2016). Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources*, 16, 1136–1146. <https://doi.org/10.1111/1755-0998.12570>.
- Trontin, J.-F., Teyssier, C., Avila, C., Debille, S., Le Matté, C., Lesage-Descauses, M., et al. (2014). Molecular phenotyping of Maritime pine somatic plants transformed with an RNAi construct targeting cinnamyl alcohol. In *Proceedings of the third international conference of the IUFRO unit 2.09.02: Somatic embryogenesis and other Vegetative propagation technologies* (p. 140). Woody Plant Production Integrating Genetic and Vegetative Propagation Technologies. <https://doi.org/10.13140/RG.2.1.1036.7767>.
- Tsubomura, M., Kurita, M., & Watanabe, A. (2016). Determination of male strobilus developmental stages by cytological and gene expression analyses in Japanese cedar (*Cryptomeria japonica*). *Tree Physiology*, 36, 653–666. <https://doi.org/10.1093/treephys/tpw001>.
- Uddenberg, D., Abrahamsson, M., & von Arnold, S. (2016). Overexpression of *PaHAP3A* stimulates differentiation of ectopic embryos from maturing somatic embryos of Norway spruce. *Tree Genetics & Genomes*, 12, 18. <https://doi.org/10.1007/s11295-016-0974-2>.

- Uddenberg, D., Akhter, S., Ramachandran, P., Sundström, J. F., & Carlsbecker, A. (2015). Sequenced genomes and rapidly emerging technologies pave the way for conifer evolutionary developmental biology. *Frontiers in Plant Science*, 6, 970. <https://doi.org/10.3389/fpls.2015.00970>.
- Uddenberg, D., Reimegård, J., Clapham, D., Almqvist, C., von Arnold, S., Emanuelsson, O., et al. (2013). Early cone setting in *Picea abies* *acrocona* is associated with increased transcriptional activity of a MADS box transcription factor. *Plant Physiology*, 161, 813–823. <https://doi.org/10.1104/pp.112.207746>.
- Ujino-Ihara, T., Taguchi, Y., Yoshimura, K., & Tsumura, Y. (2003). Analysis of expressed sequence tags derived from developing seed and pollen cones of *Cryptomeria japonica*. *Plant Biology*, 5, 600–607. <https://doi.org/10.1055/s-2003-44690>.
- Ujino-Ihara, T., Yoshimura, K., Ugawa, Y., Yoshimaru, H., Nagasaka, K., & Tsumura, Y. (2000). Expression analysis of ESTs derived from the inner bark of *Cryptomeria japonica*. *Plant Molecular Biology*, 43, 451–457. <https://doi.org/10.1023/A:1006492103063>.
- Valledor, L., Pascual, J., Meijón, M., Escandón, M., & Cañal, M. J. (2015). Conserved epigenetic mechanisms could play a key role in regulation of photosynthesis and development-related genes during needle development of *Pinus radiata*. *PLoS One*, 10, e0126405. <https://doi.org/10.1371/journal.pone.0126405>.
- Van Zyl, L., von Arnold, S., Chen, Y., Egertsdotter, U., Mackay, J., Sederoff, R., et al. (2002). Heterologous array analysis in Pinaceae: Hybridization of *Pinus taeda* cDNA arrays with cDNA from needles and embryogenic cultures of *P. taeda*, *P. sylvestris* or *Picea abies*. *Comparative and Functional Genomics*, 3, 306–318. <https://doi.org/10.1002/cfg.199>.
- Villalobos, D. P., Díaz-Moreno, S. M., Said, E. S. S., Cañas, R. A., Osuna, D., Van Kerckhoven, S. H. E., et al. (2012). Reprogramming of gene expression during compression wood formation in pine: Coordinated modulation of S-adenosylmethionine, lignin and lignan related genes. *BMC Plant Biology*, 12, 100. <https://doi.org/10.1186/1471-2229-12-100>.
- Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnology*, 32, 223–226. <https://doi.org/10.1038/nbt.2839>.
- Wachowiak, W., Trivedi, U., Perry, A., & Cavers, S. (2015). Comparative transcriptomics of a complex of four European pine species. *BMC Genomics*, 16, 234. <https://doi.org/10.1186/s12864-015-1401-z>.
- Wadenbäck, J., Von Arnold, S., Egertsdotter, U., Walter, M. H., Grima-Pettenati, J., Goffner, D., et al. (2008). Lignin biosynthesis in transgenic Norway spruce plants harboring an antisense construct for cinnamoyl CoA reductase (CCR). *Transgenic Research*, 17, 379–392. <https://doi.org/10.1007/s11248-007-9113-z>.
- Wagner, A., Donaldson, L., Kim, H., Phillips, L., Flint, H., Steward, D., et al. (2009). Suppression of 4-coumarate-CoA ligase in the coniferous gymnosperm *Pinus radiata*. *Plant Physiology*, 149, 370–383. <https://doi.org/10.1104/pp.108.125765>.
- Wagner, A., Phillips, L., Narayan, R. D., Moody, J. M., & Geddes, B. (2005). Gene silencing studies in the gymnosperm species *Pinus radiata*. *Plant Cell Reports*, 24, 95–102. <https://doi.org/10.1007/s00299-004-0911-y>.
- Wagner, A., Ralph, J., Akiyama, T., Flint, H., Phillips, L., Torr, K., et al. (2007). Exploring lignification in conifers by silencing hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase in *Pinus radiata*. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 11856–11861. <https://doi.org/10.1073/pnas.0701428104>.
- Wagner, A., Tobimatsu, Y., Goeminne, G., Phillips, L., Flint, H., Steward, D., et al. (2013). Suppression of CCR impacts metabolite profile and cell wall composition in *Pinus radiata* tracheary elements. *Plant Molecular Biology*, 81, 105–117. <https://doi.org/10.1007/s11103-012-9985-z>.

- Wagner, A., Tobimatsu, Y., Phillips, L., Flint, H., Geddes, B., Lu, F., et al. (2015). Syringyl lignin production in conifers: Proof of concept in a Pine tracheary element system. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 6218–6223. <https://doi.org/10.1073/pnas.1411926112>.
- Wagner, A., Tobimatsu, Y., Phillips, L., Flint, H., Torr, K., Donaldson, L., et al. (2011). CCoAOMT suppression modifies lignin composition in *Pinus radiata*. *Plant Journal*, 67, 119–129. <https://doi.org/10.1111/j.1365-313X.2011.04580.x>.
- Wan, T., Liu, Z. M., Li, L. F., Leitch, A. R., Leitch, I. J., Lohaus, R., et al. (2018). A genome for gnepophytes and early evolution of seed plants. *Nature Plants*, 4, 82–89. <https://doi.org/10.1038/s41477-017-0097-2>.
- Wan, L. C., Zhang, H., Lu, S., Zhang, L., Qiu, Z., Zhao, Y., et al. (2012). Transcriptome-wide identification and characterization of miRNAs from *Pinus densata*. *BMC Genomics*, 13, 132. <https://doi.org/10.1186/1471-2164-13-132>.
- Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology*, 32, 947–951. <https://doi.org/10.1038/nbt.2969>.
- Warren, R. L., Keeling, C. I., Yuen, M. M. S., Raymond, A., Taylor, G. A., Vandervalk, B. P., et al. (2015). Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *Plant Journal*, 83, 189–212. <https://doi.org/10.1111/tpj.12886>.
- Watanabe, Y. (2011). Overview of plant RNAi. In H. Kodama & A. Komamine (Eds.), Vol. 744. *Methods in molecular biology* (Clifton, N.J.): (pp. 1–11). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-61779-123-9_1.
- Wegrzyn, J. L., Lee, J. M., Tearse, B. R., & Neale, D. B. (2008). TreeGenes: A forest tree genome database. *International Journal of Plant Genomics*, 2008, 412875. <https://doi.org/10.1155/2008/412875>.
- Wegrzyn, J. L., Liechty, J. D., Stevens, K. A., Wu, L. S., Loopstra, C. A., Vasquez-Gross, H. A., et al. (2014). Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics*, 196, 891–909. <https://doi.org/10.1534/genetics.113.159996>.
- West, P. W. (2014). *Growing plantation forests. Growing plantation forests (second)*. New Delhi, India: Springer International Publishing. <https://doi.org/10.1007/978-3-319-01827-0>.
- Whetten, R., Sun, Y. H., Zhang, Y., & Sederoff, R. (2001). Functional genomics and cell wall biosynthesis in loblolly pine. *Plant Molecular Biology*, 47, 275–291. <https://doi.org/10.1023/A:1010652003395>.
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America*, 111, E4859–E4868. <https://doi.org/10.1073/pnas.1323926111>.
- Xie, W., Huang, A., Li, H., Feng, L., Zhang, F., & Guo, W. (2017). Identification and comparative analysis of microRNAs in *Pinus massoniana* infected by *Bursaphelenchus xylophilus*. *Plant Growth Regulation*, 83, 223–232. <https://doi.org/10.1007/s10725-016-0221-8>.
- Yakovlev, I. A., & Fossdal, C. G. (2017). *In silico* analysis of small RNAs suggest roles for novel and conserved miRNAs in the formation of epigenetic memory in somatic embryos of Norway spruce. *Frontiers in Physiology*, 8, 674. <https://doi.org/10.3389/fphys.2017.00674>.
- Yakovlev, I. A., Fossdal, C. G., Johnsen, Ø., Junttila, O., & Skroppa, T. (2006). Analysis of gene expression during bud burst initiation in Norway spruce via ESTs from subtracted cDNA libraries. *Tree Genetics & Genomes*, 2, 39–52. <https://doi.org/10.1007/s11295-005-0031-z>.

- Yang, I. S., & Kim, S. (2015). Analysis of whole transcriptome sequencing data: Workflow and software. *Genomics & Informatics*, 13, 119–125. <https://doi.org/10.5808/GI.2015.13.4.119>.
- Yeaman, S., Hodgins, K. A., Suren, H., Nurkowski, K. A., Rieseberg, L. H., Holliday, J. A., et al. (2014). 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*). *The New Phytologist*, 203, 578–591. <https://doi.org/10.1111/nph.12819>.
- Zhang, J., Wu, T., Li, L., Han, S., Li, X., Zhang, S., et al. (2013). Dynamic expression of small RNA populations in larch (*Larix leptolepis*). *Planta*, 237, 89–101. <https://doi.org/10.1007/s00425-012-1753-4>.
- Zhu, T., Moschou, P. N., Alvarez, J. M., Sohlberg, J. J., & Arnold, S. (2016). *WUSCHEL-RELATED HOMEODOMAIN 2* is important for protoderm and suspensor development in the gymnosperm Norway spruce. *BMC Plant Biology*, 16, 19. <https://doi.org/10.1186/s12870-016-0706-7>.
- Zhu, T., Moschou, P. N., Alvarez, J. M., Sohlberg, J. J., & Von Arnold, S. (2014). *WUSCHEL-RELATED HOMEODOMAIN 8/9* is important for proper embryo patterning in the gymnosperm Norway spruce. *Journal of Experimental Botany*, 65, 6543–6552. <https://doi.org/10.1093/jxb/eru371>.
- Zimin, A., Stevens, K. A., Crepeau, M. W., Holtz-Morris, A., Koriabine, M., Marçais, G., et al. (2014). Sequencing and assembly of the 22-Gb loblolly pine genome. *Genetics*, 196, 875–890. <https://doi.org/10.1534/genetics.113.159715>.
- Zimin, A. V., Stevens, K. A., Crepeau, M. W., Puiu, D., Wegrzyn, J. L., Yorke, J. A., et al. (2017). An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. *GigaScience*, 6, 1–4. <https://doi.org/10.1093/gigascience/giw016>.