UNIVERSITÀ DEGLI STUDI DI UDINE

SISSA

JOINT PHD PROGRAMME IN MOLECULAR BIOLOGY

Genome assembly of *Populus nigra var. italica* and genetic basis of columnar growth habit

A Ph.D. dissertation Work

By

Bhumika Dubay

To fulfil the grade requirement of

Joint Ph.D. program in molecular biology
(JuMBOO)

Supervisor
Prof. Michele Morgante

Table of Contents

1 Acronyms

AD: Allelic Depth

BAM: Binary Alignment/Map

BP: Base Pairs

ChIP-Seq: Chromatin Immunoprecipitation Sequencing

CNV: Copy Number Variant

DHH: DNA transposon Helitron

DIRS: Dictyostelium Intermediate Repeat Sequence

DOC: Depth Of Coverage

DP: Filtered Depth of individual

DTA: DNA transposon TIR hAT

DTC: DNA transposon TIR CACTA

DTH: DNA transposon TIR PIF-Harbinger

DTM: DNA transposon TIR Mutator

DTT: DNA transposon

TIR Tc1-Mariner

DTX: DNA transposon TIR unknown

DXX: DNA transposon unknown unknown

EU: European Union

F1: First filial generation of offspring

Gb: Giga base pairs

GO: Gene Ontology

IBD: Identity by descent

IBS: Identity By State

L50: No. of sequence with length greater than N50 value

LINE: Long Interspersed Nuclear Element

LTR: Long Terminal Repeat

MAF: Minor Allele Frequency

Mb; Mbp: Mega base pairs

MYA: Million Years Ago

N50: The sequence length of the shortest contig at 50% of the total assembly length

NGS: Next Generation Sequencing

NMD: Nonsense-mediated decay

OLA: Overlap layout consensus

ONT: Oxford Nanopore Technology

PCA: Principle Component Analysis

PE: Paired End

RIL: Retrotransposon LINE L1

RLC: Retrotransposon LTR Copia

RLG: Retrotransposon LTR Gypsy

RLR: Retrotransposon LTR Retrovirus

RLX: Retrotransposon LTR unknown

RXX: Retrotransposon unknown unknown

RTA: Representative Transcript Assembly

SNP: Single Nucleotide Polymorphism

SPET: Single Primer Enrichment Technology

SRA: Short Read Archive

SV: Structural Variant

TE: Transposable Element

TIR: Terminal Inverted Repeat

V: Version

VCF: Variant Call Format

XXX: Unclassified transposable element

2 Abstract

This doctoral dissertation presents a comprehensive genetic analysis of *Populus nigra var. italica*, a forest tree variety renowned for its distinctive columnar growth habit. This unique variety holds significance for its ornamental appeal and its efficacy as a windbreaker along streets and riversides, particularly in Europe. Over the past century, selective breeding has led to artificial propagation, induced mutations, and hybridization, blurring its natural habitat and evolutionary history.

The study highlights the wide distribution of the genuine italica clone in Europe. Furthermore, it introduces a reference genome assembly for *P. nigra var. italica*, emphasizing its superior quality compared to related species. The role of the transposable elements in shaping diversity and adaptation is also investigated within the genome.

To facilitate population-level analysis, a hybrid genome assembly approach, combining short-read (10X) and long-read (ONT) sequencing technologies with Hi-C scaffolding, yields a 417Mbp genome with an N50 of 9.5Mbp, along with 40,988 gene models and 34.76% transposable element content. Leveraging this genome, the research identifies informative variants for population analysis and the causative factors for columnar growth. Comparative population-level analysis involving native and non-native *P. nigra var. italica* trees and other *Populus* species uncovers two distinct introductions of italica in Europe and identifies genetic relationships among accessions.

The investigation into the columnar phenotype reveals that the tiller angle control gene, responsible for shoot-to-branch angle regulation, plays a role in the columnar trait's mechanism in various species, including poplar hybrids. It is found that a premature stop codon in italica trees, present in San Giorgio, triggers nonsense-mediated RNA decay of the gene, resulting in more vertical branches.

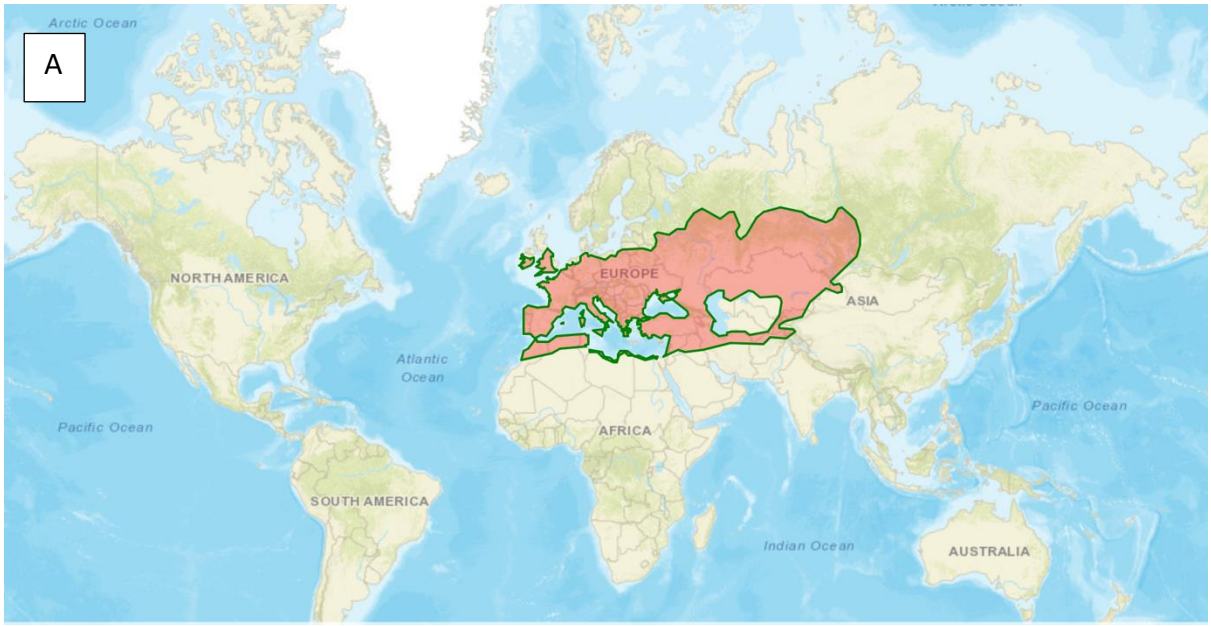
This research significantly challenges established assumptions about the history, origin, and European distribution of Lombardy poplar, revealing intricate evolutionary patterns. It elucidates the genetic basis of the columnar growth habit, identifying the gene and multiple mutations responsible for this trait. These findings have implications for breeding and genetic

engineering in Lombardy poplar and related species, contributing to our understanding of genetic diversity, species dynamics, and human breeding impacts. The shared genome assembly serves as a valuable resource for comparative genomics in tree species, enhancing our knowledge of plant diversity, adaptation, and evolution.

3 Introduction

3.1 The geographical range of natural distribution of the species *P. nigra*

The species *P. nigra* is naturally distributed across Europe and extends into central and Western Asia and northern Africa. It has also been introduced in some parts of America and Australia. In Australia, *P. nigra* is frequently found near watercourses in agricultural regions and has been observed in undisturbed riparian vegetation (Carr & Walsh, 1996). Although considered native in Europe, the gene pools of vulnerable *P. nigra* populations face pollution due to frequent hybridization, leading to the formation of bottomlands forests in the area (Ayanz et al., 2016). In northern America, the species was introduced from the Mediterranean region to control wind speed (*Populus nigra*, Go Botany, n.d.). Figure 1 provides a map illustrating the current distribution of the tree (Carr & Walsh, 1996; Ayanz et al., 2016).



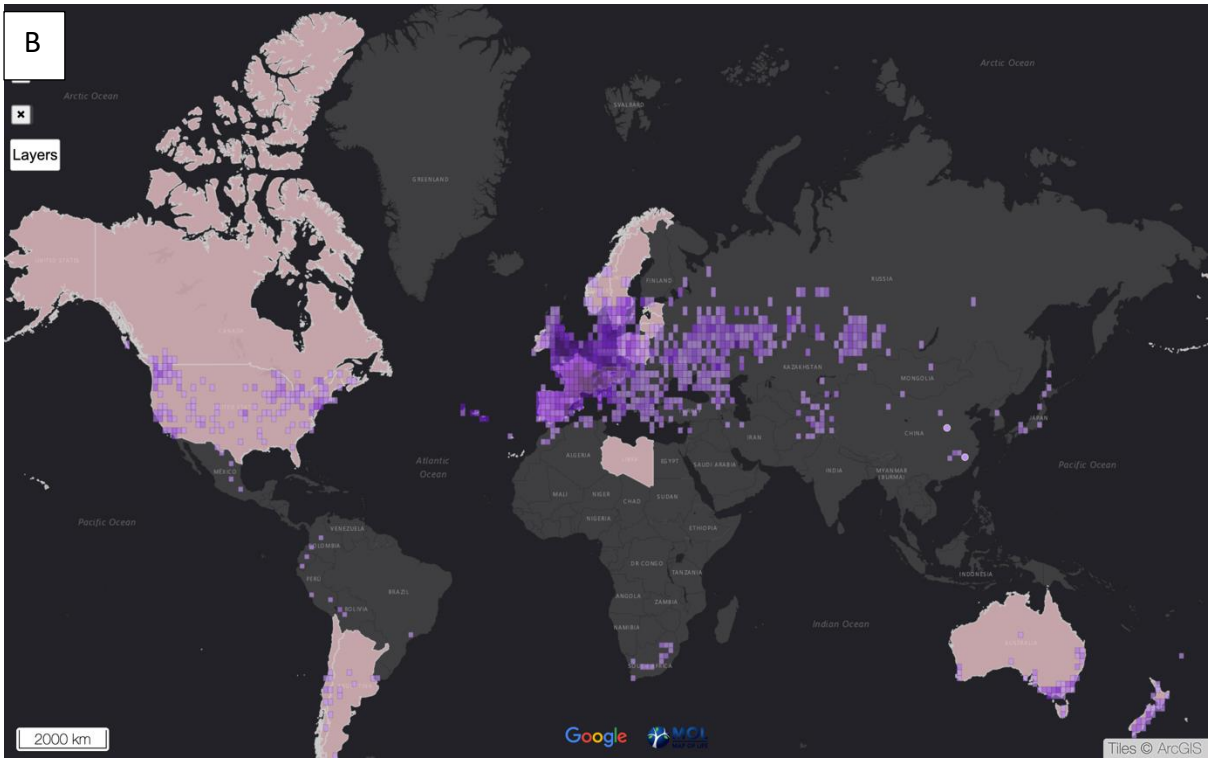


Figure 1: A: A map of the natural (native) range boundaries (highlighted in green) of the species *Populus nigra*, indicating its native presence in Europe, the Middle East, and some parts of North Asia regions (*Populus Nigra* - Black Poplar Range Map, n.d.). **B:** A map illustrating the current global distribution of *P. nigra* (Figures reference: Map of Life - Species map, n.d.). The graphic portrays the abundance of *P. nigra* trees throughout Europe, with additional occurrences in Asia, Australia, and North America. The purple color palette signifies the tree's presence, with darker shades indicating a higher density within a 10km radius of the point. The pink color denotes countries where *P. nigra* is declared an alien invasive species (e.g., North America, North Africa, and Australia).

3.2 Geographical structure of genetic diversity in the species *P. nigra* in Europe

Populus nigra is found in various geographical settings and has been categorized into various subspecies, varieties, and hybrids. Two major subspecies found in Europe are *nigra* L. and *betulifolia*. Subsp. *nigra* L. is the typical *P. nigra* found in most parts of Europe. Subsp. *betulifolia*, on the other hand, is found in northern and western Europe and grows in drier habitats such as hillsides and uplands (Rushforth, 1999).

It is understood that *Populus nigra* L. multiplies both vegetatively and sexually in spontaneous populations throughout Europe. The geographic range of clonal clusters in natural populations is relatively small. According to research on 17 populations from 11 river basins in seven European catchment systems (the Danube, Ebro, Elbe, Po, Rhine, Rhone, and Usk), genetically similar or even identical populations of *P. nigra* L. are regularly found in geographical proximity to one another. The research demonstrated that populations from the Danube catchment in Austria were more genetically like the Elbe catchment population in the Czech Republic than the Danube catchment population in Ukraine, even though they were on different riverbanks, indicating local breeding of the trees (Smulders et al., 2008). Several breeding initiatives for the tree have resulted in many clonal individuals that are dispersed over Europe and used in short cycle plantation forestry.

The female *P. nigra* tree prefers to mate with a restrictive number of male partners and is known to mate more frequently when non-*nigra* poplars (species or hybrids) are nearby in the spontaneous populations spread throughout Europe. Additionally, it is thought that this introgression is counter-selected when the seedlings are young. The male cultivar can hybridize with natural populations nearby, but in some circumstances (Belgium, UK), non-synchrony of flowering may reduce the threat (EUFORGEN Technical Guidelines for Genetic Conservation and Use for Black Poplar (*Populus Nigra*), 2003). This has resulted in frequent hybridization between *P. nigra* and other Poplar species in Europe, polluting the true *nigra* tree pool (Broeck & Božič, 2010).

The spreading of columnar *P. nigra* is primarily a result of vegetative propagation, where cuttings from parent trees are used to produce new trees with the same genetic makeup. As a result, the genetic diversity of cultivated columnar *P. nigra* is likely to be limited (Bannoud & Bellini, 2021). However, natural hybridization with other *Populus nigra* varieties and other poplar species may contribute to some genetic diversity.

3.3 Threats to the native populations of *P. nigra*

Recently, *P. nigra* spontaneous populations have encountered habitat loss and fragmentation, as well as invasive species competition. The three major threats to the species causing these issues are:

A. The alteration of riparian ecosystems throughout the species' distribution area by human activities.
B. Overexploitation of autochthonous black poplar resources and planting faster-growing hybrid poplars to replace those populations.
C. Introgression from cultivated clones and other *Populus* species.

Conservation initiatives are underway to preserve these populations and restore degraded riparian habitats, both crucial for maintaining the ecological and cultural qualities associated with this significant tree species. These initiatives include, but are not limited to, habitat restoration, genetic conservation, population monitoring, and education and outreach (Vanden, et al., 2021; Gaudet, 2006).

3.4 Varieties of columnar *P. nigra* and its geographical distribution

Within the scope of this doctoral dissertation, the focus lies on poplar trees distinguished by their fastigiate crowns, an architectural trait achieved through branches tightly adhering to the trunk. Commonly referred to as cypress poplars, these arboreal entities are exemplified by specific variations, each with its unique characteristics:

Firstly, we encounter *Populus nigra var. Italica* Duroi, colloquially known as Lombardy poplar. The historical context surrounding its emergence in the early 18th century remains debatable, with propositions attributing its origin to various locales, including the Mediterranean and Persian regions. This cultivar garners attention due to its predominantly male representation.

Secondly, the discourse shifts to *Populus nigra* var. *afghanica* (syn. *P. nigra* var. *thevestina*; Bean & Clarke, 1989), a prominently distributed variety across South-West Asia. It is entwined in taxonomic discussions, perceived by some as a distinct *P. afghanica* species and by others as a cultivar stemming from *P. nigra*, distinguished by its female population.

Progressing forward, we delve into poplars endowed with columnar or fastigiate crowns, specifically the hybrid clone ensemble called Plantierensis. Originating from the Plantières nursery in France during 1884, this assembly is a product of the amalgamation between *P. nigra* var. *Italica* and *P. nigra* betulifolia. Notably, these trees find widespread cultivation across Britain and Ireland under the epithet 'Lombardy Poplar'. The male counterpart, termed "Manchester Poplar," a subspecies derived from *P. Betulifolia*, thrives within the northwestern regions of England. Concomitantly, an enigmatic female clone labeled "Gigantea" further enriches this intricate panorama of fastigiate-crowned poplar diversity (Frison, 2014).

3.5 History and distribution of *Populus nigra var. italica*

P. nigra var. *Italica* Duroi is the oldest variety described for the *Populus nigra* species, and its origin is uncertain. According to an article on *P. nigra* (*Populus nigra - Trees and Shrubs Online*, n.d.), it was initially considered that the variety Italica originated somewhere in central Asia and was brought from there to the rest of the world through breeding and propagation. However, this theory was later proven wrong as further analysis revealed that it was a variety Afghanica with an Asian origin and was confused for Italica because of its columnar growth habits. According to some old scripts, the species was first introduced to the Po Valley, Italy, in the 17th century (Bean & Clarke, 1989) and was commonly referred to as Lombardy poplar from thereon. The actual genetic origin of the tree, however, remains a mystery and needs

more research-based evidence to make a footprint on its evolution. This research focuses on understanding how the tree emerged and propagated across Europe.

The variety Italica is cultivated in lowland areas, river valleys, pond sides, forest boundaries, and roadsides; its pyramidal shape makes it ideal for these settings. As for the geographical distribution, the tree is extensively cultivated worldwide due to its tall, columnar shape and fast growth, following its introduction into numerous countries for ornamental and practical purposes. It is commonly found across Europe, North America, Australia, New Zealand, South America, and various parts of Asia. However, its cultivation has also raised concerns about invasiveness in some regions.

3.6 Properties and use of columnar *P. nigra (var. italica)*

The columnar *P. nigra* (*Populus nigra 'Italica'*) variety is renowned for its narrow crown. It prefers moist soils and total exposure but can tolerate some drought and urban pollution. It is a fast-growing species that can be propagated vegetatively from cuttings (Figure 2). It can be susceptible to diseases and pests such as cankers, leaf rust, and aphids, as with all *Populus nigra* varieties (Ayanz et al., 2016).

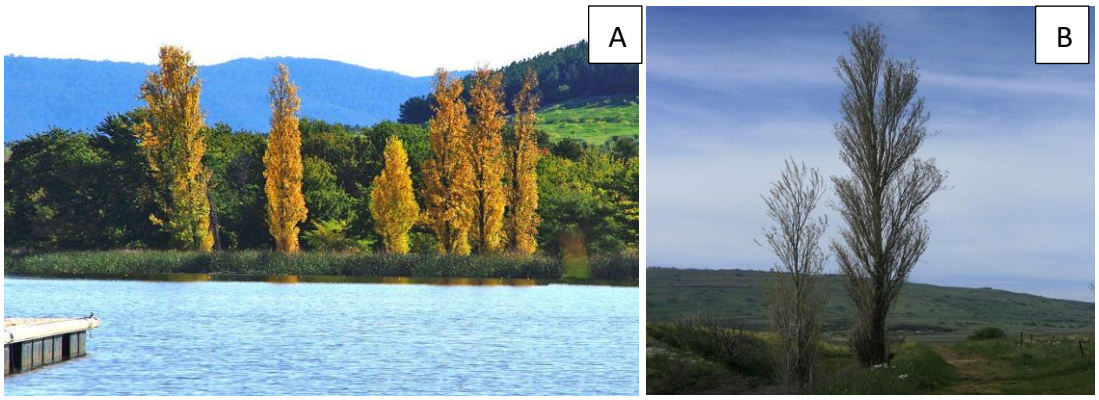


Figure 2: Two pictures of the *Populus nigra var. italica*, also known as the columnar *P. nigra* tree. Image A describes the tree beautiful and tall, which is why people wanted to use it as an ornament and as windbreaker on roadsides. The upright columnar growth habit that results in a wide crown tree is shown in Image B. These columnar habits are identifier of the italica clone (File: *Populus nigra Italica*-Golan Heights.jpg-Wikimedia Commons, 2013)

Columnar *P. nigra* are fast-growing trees, and their physical characteristics (as described above) make them suitable for growing in nurseries and all temperature settings worldwide; hence, it is used as an ornamental tree. It is one of the 27 emerging alien invasive species of the alpine of southern Africa (South Africa, Lesotho) (Carbutt, 2012). The tree is suitable for short rotation coppice. It usually has speedy juvenile growth and tends to regrow and recover from physical damages rather quickly, i.e., it possesses high totipotency (Lewandowski, 2017).

Similarly, the columnar growth habits of the tree are especially advantageous as they result in a high wood yield. We can plant more trees in the same area, resulting in a high quantity of timber. One benefit of pyramidal development may be the more effective light penetration into lower canopy layers, which boosts the plant's capacity for photosynthetic activity, making trees fast-growing (Sakamoto et al., 2006).

3.7 Hybrid genome assembly and annotation

Exploring the genetic features and population-level understanding of the columnar *P. nigra* is crucial for enhancing our knowledge of the species. A pivotal aspect of genetic analysis involves having a species and variety-specific reference genome, which can be achieved through a hybrid genome assembly technique. This method involves constructing a contiguous genome by merging different sequencing strategies, typically combining error-prone long-read assembly with short-read assembly to correct sequences and enhance overall assembly quality while reducing costs (Koren et al., 2012). To mitigate the expenses associated with high-coverage long-read sequencing, our genome assembly utilized a hybrid technique, concurrently incorporating 10X linked reads as short reads to phase resolve our diploid genome. Additionally, we employws merging and consensus generation tools like quickmerge to prevent redundancy in the genome.

A genome assembly is usually accompanied by identifying and characterizing its coding and non-coding regions. This process is called genome annotation. The features that can be identified with this process include the genes, promoters, regulatory regions, transposable elements, and other repeats. Various pipelines are publicly available to make this process easier and user-friendly, e.g., MAKER, GAAP, DFAST, and OmixBox (Abril & Castellano, 2019).

The structural annotation uses a homology-based search employing relevant publicly available databases, e.g., Entrez Gene. This method is suitable for a superficial analysis. Still, for a detailed downstream analysis, which may require precise start and end coordinates, it is ideal to integrate different machine learning-based approaches to achieve the required level of accuracy (Pevsner, 2009).

Most eukaryotic organisms are usually high in repeat content, which makes it essential to identify these elements to avoid misannotation of genomic fragments and remove assembly bias due to hard-to-assemble repeat regions. The de-novo approach tends to identify transposable elements (TE) without using a guide sequence and with the help of either features like long terminal repeats, inversion sequences, the copy number of the fragments, or using machine learning algorithms and k-Mer-based models. Some tools for de-novo TE identification are EDTA (Ou et al., 2019), REPET (Flutre et al., 2011), and RepeatMasker (Smit et al., n.d.). The de-novo method's limitations lie in mischaracterizing similar-looking elements and false-positive discovery (Goubert et al., 2022).

3.8 Population Genomics Studies

Population genomics is a sub-field of genomics that seeks to comprehend the genomic variations within and between the populations of the organism of interest. This is a useful approach to study evolution genome-wide (Wright, 1931). With the development of DNA sequencing tools, it is now simpler to locate and target specific markers within the genome that might be related to characteristics of interest, making population genomics studies easier. It takes many mathematical models to comprehend and understand a more significant amount of population-level data.

3.8.1 Haplotype Sharing Analysis

A haplotype is a multi-locus haploid genotype. Identifying shared ancestry between individuals in a population requires identifying the segments of shared haplotypes. In population genomics, haplotype-sharing analysis is usually done to identify the regions in the

genome that are shared among individuals due to their being inherited from common ancestors (Howard et al., 2021). Based on the degree of haplotype sharing, the analysis explains the degree of relatedness between individuals in the population.

The critical concept in haplotype-sharing analysis is identity by descent (IBD). IBD describes whether two individuals or groups share the same haplotype and whether it is due to common ancestry or by chance. IBD analysis helps determine the relatedness between the members of a common ancestry (Browning, 2008).

3.8.2 Identification Of Causative Mutations

Identifying the mutation causing a specific phenotype is the final step in understanding a trait's genetic control. The path to identifying a causative mutation can follow various approaches such as linkage mapping, genome-wide association studies, and the analysis of functional gene candidates.

Linkage analysis works on the principle that any sequence variation sufficiently near the gene or the mutation causing a trait variation tends to be inherited together in all the individuals of a segregating family with the phenotypic variation for that trait. By examining how the genetic markers are inherited in individuals with the trait, the possible lesion causing the trait can be identified. On the other hand, association studies focus on allelic frequencies of genetic markers in individuals of large natural populations with and without the trait separated by several generations (Cooper, et al., 2013). However, in the presence of linkage disequilibrium between the sequence variation and the causal mutation, if the frequency of a genetic variant is higher when the trait is present, it is associated with the trait.

A candidate gene approach identifies a causative mutation responsible for a particular phenotype or trait by focusing on a specific gene or set of genes known or suspected to be involved in the biological pathways underlying the phenotype or trait of interest (Kwon & Goate, 2000). This approach assumes that a particular gene or set of genes is likely to be associated with the phenotype or trait of interest based on previous studies, known biological

functions, or association studies (Zhu & Zhao, 2007). Once the candidate genes are identified, they can be sequenced or genotyped in individuals with the phenotype or trait of interest and compared to individuals without the phenotype or trait to identify any genetic variations associated with the phenotype or trait.

One of the advantages of the candidate gene approach is that it can be a cost-effective and efficient method of identifying causative mutations, as it narrows down the search to a specific set of genes or pathways. However, one of the limitations of this approach is that it requires prior knowledge or hypotheses about the genes or pathways involved in the phenotype or trait of interest, which may not always be available or accurate. The candidate gene approach may also miss causative mutations outside the candidate genes or pathways. Therefore, it is often combined with other approaches, such as genome-wide association studies or whole-genome sequencing, to identify causative mutations comprehensively.

3.8.3 Genes involved in columnar habits of the tree

The columnar habits in plant growth are frequently associated with the genes that control the shoot-to-stem angle. The two genes identified as involved in regulating this angle are LAZY1 and TAC1. However, they have diverse roles and activities.

The LAZY1 gene regulates plant growth orientation, specifically lateral shoot growth and adventitious root production. Downregulation of LAZY1 can cause a "lazy" or "prostrate" phenotype, in which lateral shoot growth is inhibited, resulting in a more horizontal growth habit (Taniguchi et al., 2017). TAC1 (TILLER ANGLE CONTROL 1), on the other hand, is involved in regulating the angle of tiller growth in grasses such as rice and wheat. Tillers are lateral shoots that grow from the plant's base and contribute to overall plant yield. The TAC1 gene regulates the angle of tiller growth. Downregulation of this gene leads to a more upright plant architecture and an increased angle of tiller growth (Dardick et al., 2013). The gene has also previously been directly associated with pyramidal growth in *P. x Canescens* in a study where the homozygous knock-out of the ortholog of the TAC1 gene in the Poplar species resulted in upright growth (Fladung, 2021).

Despite their distinct roles in plant growth and architecture, the LAZY1 and TAC1 genes contribute to the general regulation of plant form and function. Understanding how these genes work and interact with other genetic and environmental factors will help us better understand columnar plant growth.

4 Thesis objectives

The main objective of this thesis is to understand the origin of true-to-type *italica* clones in *Populus nigra* var. *italica*. We would also explore the genetic properties of pyramidal P. nigra clones obtained from different European countries. This objective extends to understanding the genetic relationship among accessions classified as pyramidal and further identifying the mutation causing the distinct *italica* clone feature, i.e., columnar growth habit.

4.1 The genomic structure of *Populus*

The *Populus* species has a small genome size of about 500 Mbp (Tuskan, et al., 2006). The haploid chromosome number *n* is 19, and trees are diploid with *2n*=38. As the tree is diploid, individuals are usually highly heterozygous. *Populus* species are also rich in repetitive regions, with more than 40-50% of the genome covered by transposable elements due to a recent duplication event that affected both genic and non-genic regions. The class I LTR-Gypsy is the most common family of TEs. On average, *poplar* species contain a little less than 50,000 genes.

4.2 The origin of columnar *P. nigra*

The columnar *P. nigra*, also known as *Populus nigra 'Italica'*, is a cultivated variety of *P. nigra* (*Populus nigra*), which follows fastigiate growth and comes from Mediterranean regions of Europe. Augustine Henry discovered evidence that it originated on the banks of the River Po in Lombardy between 1700 and 1720 and spread fast by cuttings, reaching France in 1749, England in 1758, and North America in 1784 (Henry, 1914). From thereon, it was selected for its narrow, columnar form and has been widely propagated and planted in various parts of the world as a cultivated variety. However, the precise mechanisms underlying the diversification of columnar *P. nigra* remain unknown.

4.3 The factors causing columnar growth

According to some sources, the tree's narrow, upright growth pattern resulted from spontaneous mutations that Italian horticulturists carefully bred for (Zsuffa, 1974). Very little is known about the origin of the *italica* clone. From the literature, we know that a homozygous mutation in the male *P. nigra* tree was the cause of the columnar effect, creating what is botanically referred to as variant *italica* (Zsuffa, 1974). Another known fact about the columnar growth was that in the hybrid *P. × Canescens*, the tiller angle control gene was recently proved to be associated with the columnar effect through CRISPR/Cas9-mediated gene knockout (Fladung, 2021).

5 Material and Methods

5.1 Reference Genome creation

To create a draft genome of *Populus nigra* var. *italica*, we selected a male tree that belonged to the largest cluster of identical trees sampled within the EpiDiverse collection, corresponding to the most frequent genotype found in Europe. This tree is part of a clonal lineage that has been historically propagated, distributed, and planted across all of Europe (see the Results section for further details).

5.1.1 DNA extraction

To generate the reference genome (10X and HiC) through whole-genome sequencing, we employed the high molecular weight nucleotide extraction protocol (Zhang et al., 1995). In this procedure, nuclei were initially extracted from fresh-frozen plant leaves. Subsequently, high molecular weight DNA was extracted and re-suspended in 200 µl of distilled H2O, resulting in the extraction of 1800 ng total DNA.

For long-read sequencing (ONT), we utilized the Doyle & Doyle CTAB modified protocol for DNA extraction (Doyle & Doyle, 1987) to extract the sample DNA.

5.1.2 Library preparation and sequencing

The linked-read sequencing utilized the 10X Chromium platform for library preparation, followed by sequencing on the Illumina HiSeq-2500. Initially, DNA was size-selected on the Blue Pippin (Sage Science) FIGE system, targeting fragments of 40 Kb – 80 Kb. Subsequently, the library preparation protocol of 10X Genomics was employed, involving the generation of GEMs (gel beads in the emulsion) and barcoding using the 10X Genomics Chromium System. After clean-up and quality control, the library was prepared for further use. Sequencing was conducted on the Illumina HiSeq-2500 machine at IGA Technology Services facilities,

maintaining a sequence length of 250 bp per read for paired-end sequencing. Approximately ~223M reads were generated per strand, achieving an overall coverage of around 150X of the total genome. Recognizing the optimal performance of the Supernova assembler with 150 bp reads, we cropped our reads to 150 bp, resulting in a raw nominal coverage of 78X.

For Hi-C sequencing used in re-scaffolding the assembly, the Arima-HiC Kit for Plant Tissues (Arima-HiC Kit User Guide for Plant Tissues, 2019) was employed with a slight modification in the crosslinking experiment. The crosslinking mix was applied to frozen leaves (frozen-crumbled to 5mm pieces), initiating vacuum-crosslinking directly on formaldehyde-thawed leaves. Subsequent steps followed the Arima Hi-C library preparation guidelines. The final library was sequenced on the Illumina HiSeq-2500.

The Oxford Nanopore library preparation procedure involved a 1D Lambda Control Experiment (SQK-LSK109). During the DNA repair step, samples were incubated for 60 minutes at 20°C. For sequencing, 268.8 ng of the final library preparation was loaded onto the active pores of an Oxford Nanopore Flow cell, achieving a 30X coverage of raw reads. HMW DNA was used to prepare two natively ligated libraries (SLK109) and sequenced on two MinION flow cells (R9.4.1).

5.1.3 Hybrid assembly to create the draft genome sequence

We adopted a hybrid assembly approach to construct the reference genome of *P. nigra* var. italica, integrating both short linked-reads and long-reads to generate sequence contigs. The subsequent step involved genome scaffolding using Hi-C data. The reads obtained from 10X Genomics-linked read sequencing underwent processing using their assembly software, Supernova version 2.1.1. The files from the sequencer were initially converted to FASTQ format using the supernova mkfastq command. Subsequently, fastq underwent quality checking using fastQC v0.11.4. To achieve the most comprehensive assembly, reads were then trimmed to a length of 150bp using cutadupt V2.1 (as determined by BUSCO assessment, Figure 3). The assembly process utilized the supernova run command with the target set to include all available reads

(i.e., the --maxreads='all' option). The resulting output was transformed into two pseudo haplotype styles using the supernova mkoutput command.

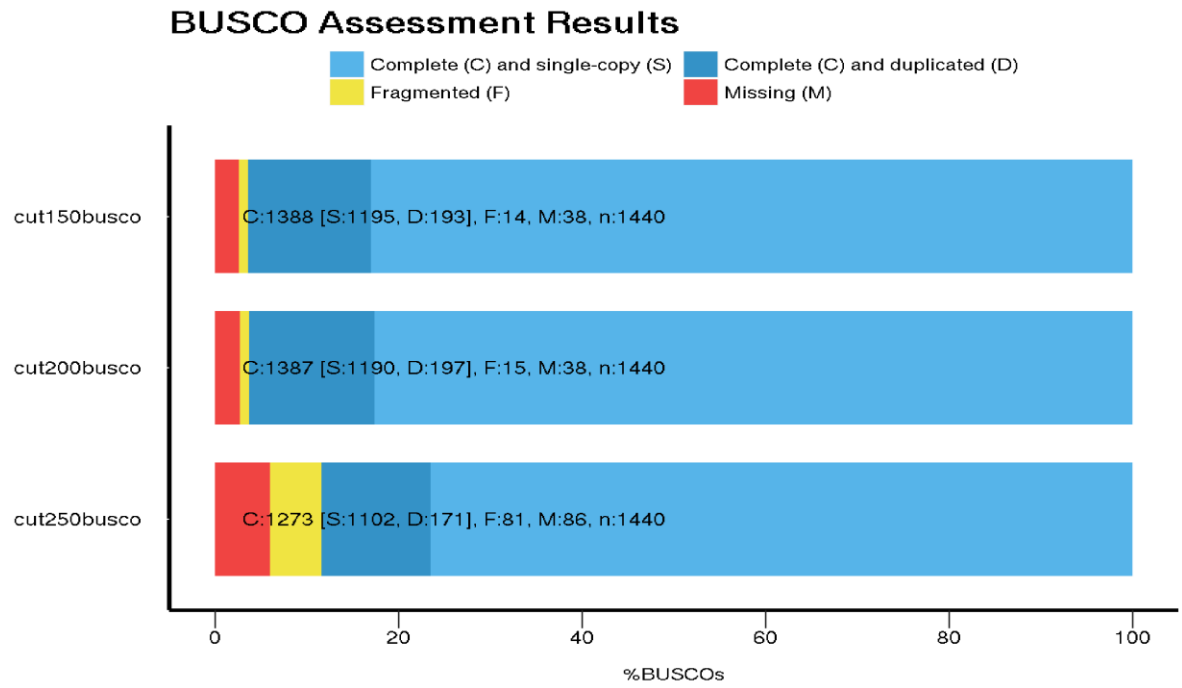


Figure 3: A chart displaying the 10X only assembly completeness measured using BUSCO statistics for the gene set with conserved orthologs. X axis represents the percentage of busco genes found in assembly carried on different read length in Y axis. N stands for the total number of genes sought, C for complete genes discovered in the assembly (S, single copy, D, double copy), F, fragmented genes, and M, missing genes. The figure highlights that the shorter reads produced better assembly and tit was most complete when the reads were cut to 150bp each.

A nanopore-only assembly was conducted using the wtdbg2 software v.2.3 (Ruan & Li, 2019), employing a low-coverage resilient algorithm with optimized parameters. The fuzzy Bruijn graph approach for long noisy reads assembly was selected. In this process, ONT sequencing type was chosen, and the estimated genome size was maintained at 423 Mb, calculated through the formula {N = n / C}, where N represents the estimated genome size, n is the total

k-mers, and C is the coverage. Crucial parameters for wtdbg2 included -k 0 -p 21 -S 1 --edge-min 3 -o S1_k0p21_e3 -K 1000.049988 -s 0.050000 -g 423000000 -X 50.000000 -L 5000. The consensus was derived using wtpoa-cns.

The long-read assembly underwent polishing with ONT reads using the minimap2 algorithm version 2.16. A second layer of error correction was executed using 10X raw reads on the long-read error-corrected assembly, employing adapter-trimmed 10X reads and Pilon version 1.23.

Simultaneously, gaps in the single haplotype assembly from 10X only reads were filled utilizing the ONT polished assembly from the previous step. This was achieved by creating windows with lengths between 10 Kb to 50 Kb for each sequence using PbJelly version 15.8.24. The blasr parameters included --minMatch 8 --minPctSimilarity 70 --bestn 1 --n candidates 20 --maxScore -500 --nproc 32 –noSplitSubreads.

Finally, the two assemblies obtained (polished ONT assembly and polished 10X assembly) underwent two rounds of merging using quickmerge v 3.1. The initial round merged the two assemblies, and the second round merged sequences that were unused in the first merging, considering the genome's diploid nature. Quickmerge employed the nucmer algorithm for mapping with -l 100 and -p, and the merging parameters were set as -hco 5.0, -c 1.5, -l 100000, and -ml 5000. The delta was filtered for -r, -q, and -l 2500. The resulting assembly was termed Version 1 assembly (V1, Figure 4).

The final merged assembly was scaffolded using Hi-C reads with SALSA version 2.2. (Ghurye et al., 2017). Hi-C reads were aligned to the V1 assembly using BWA version 0.7.17 mem algorithm with -5SP, and duplicate marking was done using the matloc tool from Phase Genomics. SALSA scaffolding utilized enzymes -e GATC, GANTC based on the library prep protocol.

In the concluding phase, manual curation addressed incorrectly scaffolded scaffolds 4 and 8 by dividing them at the site of misorientation, creating two scaffolds for each (scaffolds 4a, 4b, 8a, and 8b). Manual curation also identified chloroplast and mitochondrial sequences, with NCBI Acc. NC_037416.1 and NCBI Acc. KM091932.1 used as references. BLASTn was employed with a query coverage of 5Kb and identity and similarity percentages of 80% each. The

obtained chloroplast and mitochondrial sequences for the scaffolds (full length) with the best hits were then substituted.

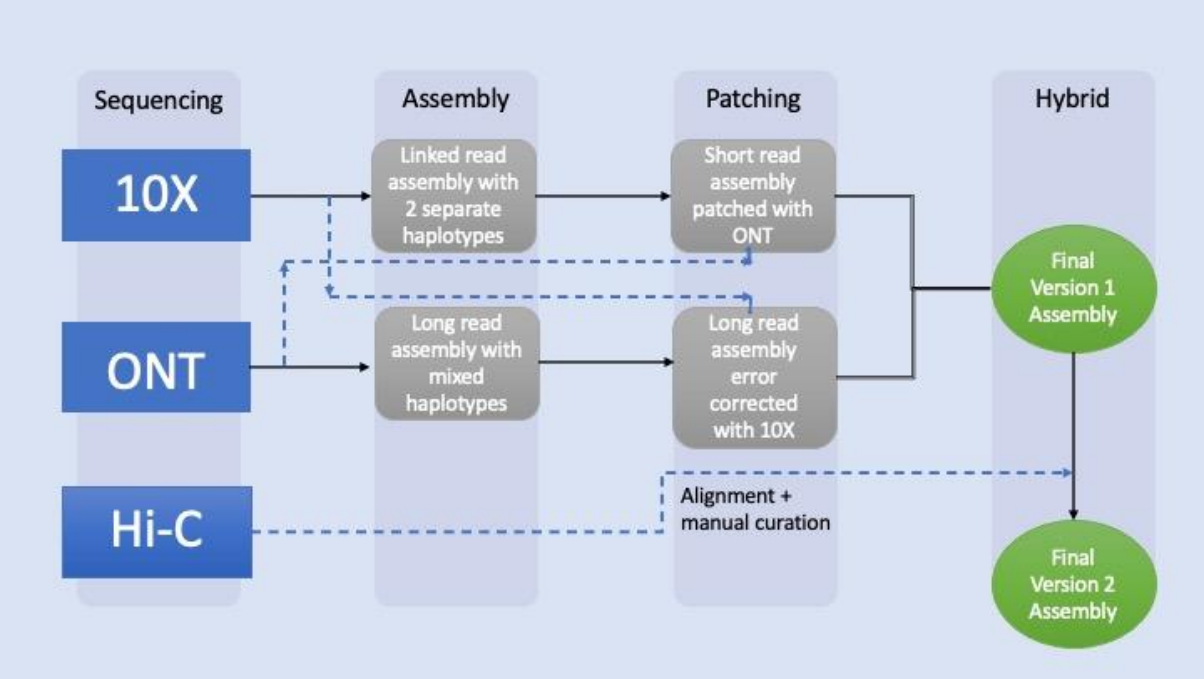


Figure 4: The flowchart depicts the pipeline used to build the reference genome through a hybrid assembly technique. To develop a phase-aware assembly, we coupled 10X linked read sequencing with ONT long read sequencing. The incorporation of Hi-C Seq data further scaffolded the assembly. This was followed by manual curation to fix major errors.

5.2 Transposable element identification and annotation

We employed Repeat Masker program version 4.0.6 to identify homologous sequences in our assembly, utilizing Eucots and Dicots plant repeat sequences from the RepBase database revision 2019 (Jurka et al., 2005). This library was also applied to mask our genome during gene annotation.

A reference-based library was collaboratively developed with Dr. Moaine Elbaidouri at the University of Perpignan, France. The library was generated by initially scanning the genome for TEs protein domains using tblastn from the Rexdb database (Neumann et al., 2019). The

extracted target regions (DNA sequences) were then clustered based on the 80/80 rule (80 percent similarity and 80 percent coverage). For each cluster, flanking regions (7 Kb left and right) were extended, followed by reverse complementing of anti-sense paralogs. Multiple alignments were performed on these sequences to produce consensus sequences, requiring a minimum of 30 percent of paralogs to share the same sequence.

The annotation pipeline EDTA V1.8.3, established for plant species, generated a library of complete TEs (Ou et al., 2019). In the final stage of identifying missing pieces, we updated the pipeline to utilize RepeatModeler2 instead of the default repeatmodeler. The pipeline operated in sensitive mode for error correction, and annotation was requested. A comprehensive evaluation of the final library was conducted.

The last de-novo library was constructed using RepeatModeler2 without reference libraries, relying on NCBI as the database source. Initially, we observed numerous false positives for helitrons in the EDTA library. To address this, we refined the EDTA library by reassessing predicted helitrons using bedtools intersect to identify EDTA-based helitron fragments in the Repeatmodeler and Rexdb-based library. Based on the annotation provided for these fragments, accurate reannotation was performed. The final library resulted from combining the polished EDTA library and repbase-based predictions using bedtools concat. Overlapping fragments within ten bp were reannotated, with preference given to the predictions made by the repbase-based library in case of class mismatches. This preference was based on the reliability of the repbase-based library, closely representing the family distribution of other poplar species (Figure 5).

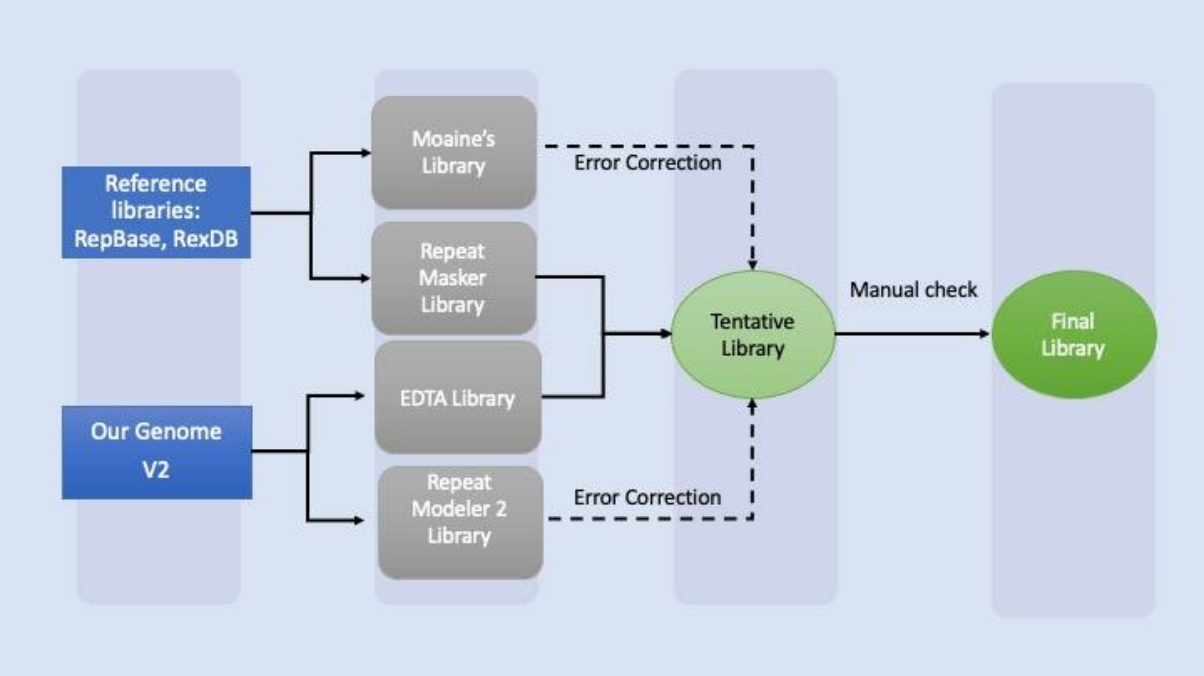


Figure 5: The flowchart describing our integrated strategy to identifying the least fragmented TEs and accurately classifying them in the shortest period. The strategy began with the creation of separate libraries. RepBase, RexDB database, EDTA, and repeatmodeler2 are four ways used. Following that, the RepBase and EDTA libraries were combined, and the categorization was revised, and error corrected using RexDB and Repeatmodeler2.

5.3 Gene structure identification and functional annotation

Gene model prediction was conducted using MAKER (Campbell et al., 2013). The initial data aiding MAKER included: 1. mRNA sequences derived from RNASeq of three tissue samples from the reference individual (obtained from an in-house poplar-based project, not published), 2. ESTs obtained from NCBI (employing all 433,547 ESTs available for Populus sp.), 3. Protein sequences covering all available rosids (utilizing all 2,324,658 protein sequences for rosids available on NCBI), and 4. mRNA sequences from NCBI for *Populus nigra* originating from 15 different tissue types, including roots, shoots, buds, xylem, and leaves (Accession PRJNA304279, WatBio project). Initially, mRNA sequences were aligned to the V2 reference using the STAR aligner with default parameters, followed by Stringtie assembler to generate

the merged gff for all mRNA sequences. The next step involved genome masking for repeat content using the RepBase library from TE annotations.

Through the amalgamation of evidence grounded in homology with plant proteins, ab initio approaches, and RNA-Seq alignments, protein-coding gene models were predicted using MAKER-P. This pipeline underwent two rounds, configuring the organism type as eukaryotic, turning off re-annotation using the MAKER-produced gff, setting the maximum DNA length at 100 Kb, and the minimum at 250 bp. MAKER round 1 produced the gff, which was then utilized to train SNAP and Augustus algorithms. The output from round one fed into the second round of MAKER to create a benchmark prediction.

Mitochondria and chloroplast annotation were accomplished using the online tool https://chlorobox.mpimp-golm.mpg.de/geseq.html

To categorize predicted genes into functional categories, biological processes, and molecular functions, the OmicxBox trial version (V 1.4 for Mac OS 10.9+) was employed. In this step, the gff3 file was input into the graphical user interface, and the BLAST2go pipeline was followed. The process comprised three major phases: 1. Utilizing the BLASTX technique, sequences were aligned to plant kingdom genes in the NCBI's nr database, with an e-value of 1.0E-2 chosen, and a similarity of more than 40% with the best hit. 2. Gene Ontology keywords were extracted from the accession of aligned genes from the previous step and mapped to the genes in the mapping step. 3. The interProScanner was used to retrieve domain/motif information in a sequence-wise manner. Corresponding GO terms were then added to the sequences and combined with previously defined GO terms.

5.4 Genetic relationships of columnar genotypes

The present population genomics investigation was conducted with the primary objective of elucidating the interrelationships existing among individuals of *Populus nigra* exhibiting a columnar growth habit, sourced from diverse regions across Europe within the framework of the EpiDiverse project. In addition to the European samples, the study encompassed

individuals originating from varied global locales, alongside non-columnar individuals serving as control groups.

5.4.1 Sample Collection, DNA extraction, and available DNA data

In the pursuit of a comprehensive population genomics investigation, samples and data were meticulously gathered from various *Populus nigra (P. nigra)* trees, constituting four distinct sets as detailed in Table 1. A subset of individuals from these sets underwent rigorous resequencing analysis.

The first set comprised 436 samples of columnar *P. nigra*, sourced from diverse locales such as botanical gardens, private gardens, urban and rural green spaces, parking lots, industrial areas, soccer field fences, roadsides, and meadows. These samples were collected across 8 European countries as part of the EpiDiverse initiative (Rodríguez et al., 2022). All specimens in this set were characterized as possessing a male genotype with columnar habits. However, due to unfavorable habitat conditions, some samples could not be definitively identified as columnar or male, herein referred to as the EpiDiverse collection (Appendix I). All samples within the EpiDiverse group were freshly collected and promptly snap-frozen with liquid nitrogen. Subsequently, these frozen samples were ground with liquid nitrogen, and DNA extraction was performed using the CTAB DNA extraction protocol.

The second set comprised 59 columnar *P. nigra* samples procured from the germplasm repository of the Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), Research Unit Foreste e Legno (FL), situated in Casale Monferrato, North-western Italy. This collection, referred to as CREA-FL (Appendix II), included specimens initially introduced from 10 countries in Europe and beyond. These samples exhibited a mix of male and female phenotypes, all displaying columnar growth habits. Dormant wood cuttings for these samples were graciously provided by CREA-FL. The primary objective of this set was to introduce the female genotype and samples from regions outside Europe into our population, alongside reference samples from a curated collection. Freshly collected apical leaves from these specimens were frozen in a -80°C freezer as opposed to liquid nitrogen. Beads-based grinding

was carried out using the TissueLyzer II from ©QIAGEN, and subsequent DNA extraction followed the CTAB-modified protocol for plant sprouts (Doyle & Doyle, 1987).

The third set comprised DNA samples from *P. nigra* trees originating from wild populations in natural sites. These trees, assumed to exhibit a typically broad crown, had been previously genotyped under the WatBio project in our laboratory (Scaglione et al., 2019). The WatBio Collection included 475 trees from Southern and Central Europe, serving as a negative control, as the columnar tree shape was not a characteristic feature observed in these specimens.

The fourth and final set encompassed a group of accessions from various *Populus* species, with whole-genome sequencing reads deposited at the NCBI Short Read Archive (Appendix IV). The primary aim of this group was to identify any distinct species or hybrids within the first three collections.

Table 1: Description of the sampling strategy to introduce maximum diversity in the analysis. Four sets of samples were used to increase the genetic diversity and reduce

Sample Set	Description	Individuals count
EpiDiverse Collection	Columnar clones collected all over Europe.	438 Samples 12 locations 8 European countries EpiDiverse project Assumed only male sex.
CREA-FL Collection	A very diverse set of columnar clones originating from all over the world.	59 Samples 20+ locations 5 European countries 4 non-European countries Both male and female sex
WatBio Collection	Individuals from natural populations covering the European range of the species	540 Samples 17 locations 5 European countries WatBio project Sex unknown
NCBI-*Populus* Collection	Short read data available for different poplar species in NCBI	28 Samples 9 Countries

The samples were selected based on PCA capturing the common variants among EpiDiverse, CREA-FL and WatBio collections for the re-sequencing whole genome analysis. The focus was capturing the maximum genetic diversity and sequencing each group representative. (Figure 12-B)

5.4.2 Targeted Genotyping and SNP calling

The library preparation for the EpiDiverse and CREA-FL collections adhered to Nugen's Allegro Targeted Genotyping protocol (Allegro® Targeted Genotyping). Subsequently, the finalized libraries were diluted in low-EDTA TE buffer and subjected to sequencing on Illumina's Nextseq 550 machine, aiming to generate 1 million reads per sample. These reads targeted 5000 genomic regions, designed in accordance with the *Populus trichocarpa* reference genome.

Base-calling was executed using the Genome Analysis Toolkit (GATK) to obtain raw fastq sequences from the EpiDiverse and CREA-FL collections. In the case of the WatBio group and various poplar species, raw fastq files were obtained from IGA's old repository (Scaglione et al., 2019) and NCBI, respectively. Notably, the targeted regions for the WatBio group, constituting a Single Primer Extension (SPET) genotyping experiment, overlapped with the 5000 regions targeted in the EpiDiverse and CREA-FL collections.

Read alignment was conducted using the Burrows-Wheeler Aligner (BWA). Preceding alignment, files underwent trimming with Cutadapt version 1.4.1, and quality assessment was performed using FastQC v0.11.4. The aligned bam files were subsequently sorted, and read groups were assigned using Picard tools v2.23.4 AddOrReplaceReadGroup. Filtering for a quality score of at least ten was applied, and the resultant files served as input for the GATK4 v4.1.8 HaplotypeCaller variant calling pipeline. The output file underwent GATK filtering to retain only Single Nucleotide Polymorphisms (SNPs) in the Variant Call Format (vcf). Further filtering removed samples with Quality by Depth (QD) < 2.0, Mapping Quality (MQ) < 40.0,

and MQRankSum < -12.5. Subsequent checks for sample and site missingness were performed, culminating in the creation of the final vcf file.

Principal Component Analysis (PCA) was carried out using the SNPrelate package from the R Bioconductor repository. Initially, the snpgdsVCF2GDS command converted the vcf file into the gds format, extracting and storing the dosage of the reference allele for both biallelic and multi-allelic SNPs, indels, and structural variants. The snpgdsPCA function was then employed to calculate principal components for each sample. Visualization of the final datasets utilized the R plot function, with distinct combinations of samples defining reliable results.

SNPs in each vcf file underwent initial filtering for coverage greater than 15 using bedtools. Subsequent filtering involved heterozygous allelic frequencies between 0.25 and 0.75 and homozygous allelic frequencies of less than 0.02 and greater than 0.98. Informative SNPs were selected if present in at least 80% of the samples in each set. Filtering also excluded SNPs falling within Transposable Element (TE) regions, as these do not accurately represent allelic diversity but rather reflect transposon events. The final filtering incorporated a minimum allelic frequency of 0.01, employing a combination of bedtools, vcftools, and custom awk scripts.

5.4.3 Resequencing and whole genome SNP calling

DNA, previously extracted from each distinct collection, was used for whole-genome resequencing experiments. The rationale for the selection of 13 samples was rooted in the strategic acquisition of one sample from each representative clonal group, as illustrated in Figure 13. The selection process prioritized samples positioned in the central regions of various clusters, ensuring the capture of the maximum conceivable variance within the dataset. Library preparation was executed using an Illumina whole-genome sequencing library prep kit, and subsequent sequencing occurred on a Hiseq 2500 machine, targeting a coverage of approximately 40X per sample.

The acquired reads underwent alignment to the reference genome assembly utilizing the BWA-mem algorithm, incorporating the 5SP option for enhanced precision. These alignments

were subjected to further refinement, filtering for quality scores exceeding 10. Read groups were subsequently assigned to Binary Alignment Map (BAM) files employing Picard tools. Genotypic calls were executed using the Genome Analysis Toolkit (GATK), following the same methodology employed in targeted genotyping as described previously.

Further curation of the Single Nucleotide Polymorphisms (SNPs) dataset involved filtering to retain exclusively biallelic positions, with subsequent exclusion of the least informative sites and samples. Heterozygous genotypes were identified when the proportions of the two alleles were falling within the range of 0.25-0.75. Homozygous reference genotypes were assigned when the alternative allele proportion was less than 0.1, while homozygous alternative genotypes were designated when the same proportion exceeded 0.9. This stringent approach ensured the retention of high-quality genotypic information, contributing to the precision of downstream analyses.

5.4.4 Data analyses

For the Principal Component Analysis (PCA), the SNPrelate package in the R programming language was employed. The calculation of pairwise genotypic distances was executed according to the following formula:

$$genD = \frac{\left((PD1 * 0.5) + PD2\right)}{(PD2 + PD1 + PD0)}$$

PD1=number of positions that differ in 1 allele,
PD2=number of positions that differ in 2 alleles,
PD0=number of positions that do not differ (i.e., identical).
Based on the pairwise genotypic distances (genD) for each individual, individuals were subsequently classified as either identical or non-identical.

The Haplotype sharing analysis was conducted following the methodology outlined in Magris et al. (2021), aiming to unveil genealogical relationships among the individuals under investigation. The reference genome underwent segmentation into genomic windows, each

encompassing 100 Kb of unmasked (i.e., non-repetitive) bases. In pairwise comparisons between individuals utilizing the Identity-By-State (IBS) threshold, as defined by Magris et al. (2021), each genomic window received a non-repetitive assignment of IBD=0, IBD=1, or IBD=2 (Table 2 delineates the IBD scoring system). An IBD=0 denoted no common haplotype, IBD=1 signified the presence of one shared haplotype, and IBD=2 indicated the presence of both shared haplotypes across the samples in a pairwise comparison for each Single Nucleotide Polymorphism (SNP) location. Genome-wide plots illustrating IBD=0, IBD=1, or IBD=2 sharing were generated employing 2402 windows situated in the 33 largest scaffolds of the Populus nigra var. italica assembly, encompassing a total of 317 Mb and corresponding to 74.5% of the estimated genome size.

Table 2: The scoring matrix for defining the Identity of State in a pair of individuals. If the individuals are homozygous for the reference, IBS is 0, i.e., both alleles are the same. If one of the two alleles is different, IBS is 1, and IBS is 2 if both alleles are different. This matrix was used for grouping the clonal individuals in our EpiDiverse and CREA-FL populations and separating them from non-clonal ones using the identity by state approach.

Site	Individual 1	Individual 2
0	AA	AA
	BB	BB
IBS 2	AB	AB
IBS 1	AB	AA or BB
	AA or BB	AB
IBS 0	AA	BB
	BB	AA

5.5 Identifying underlying mutation for columnar phenotype

The preliminary step in elucidating the causal mutation underlying the columnar phenotype involved the identification of a pertinent region or gene. A systematic literature search was undertaken, revealing three distinct genes documented in various plant species, known to influence shoot angles and directly impact overall plant architecture. These genes were identified as Co-Gene in apples, PepTAC1 in peaches, and LAZY1 in *Arabidopsis*.

Of particular significance, previous research had established a linkage between the knockout of the TAC1 gene ortholog and the manifestation of a columnar effect in other poplar species (Fladung, 2021). To facilitate a comprehensive analysis, we procured the protein sequences of Co-Gene, PepTAC1, and LAZY1. Employing the tBLASTn algorithm within the CoGe platform, a meticulous homology search was conducted.

Subsequently, the resultant genes, along with their upstream and downstream regions (excluding regions harboring Transposable Elements), were manually scrutinized using genome viewers IGV and TABLET. This investigation encompassed identical genomic areas from the 13 individuals subjected to Whole Genome Sequencing (WGS), WGS data from three non-columnar individuals sourced from the WatBio experiment, and the reference sample. This rigorous search strategy aimed to pinpoint potential candidate genes associated with the columnar phenotype, setting the stage for subsequent in-depth functional analyses.

6 Results

6.1 Construction of a reference genome assembly

6.1.1 Identification of the most common columnar genotype for *P. nigra var. italica*

The EpiDiverse project established an ex-situ collection comprising 436 columnar *P. nigra* individuals, acquired through sampling efforts conducted across eight European Union countries. These samples originated from cuttings of *Populus nigra* var. 'Italica' clones cultivated in Europe, covering latitudinal gradients from approximately 41° to 60° N and longitudinal spans from -5° to 25° E. Sampling encompassed twelve locations, reflecting diverse Köppen-Geiger climate subtypes across Europe (Peel et al., 2007; Rodríguez et al., 2022).

Utilizing Single Primer Enrichment Technology (SPET)-targeted sequencing at 5,000 loci with primers previously designed on the *P. trichocarpa* genome (Scaglione et al., 2019), a total of 2,360 Single Nucleotide Polymorphisms (SNPs) with Minor Allele Frequency (MAF) greater than 0.05 were identified within this cohort of genotypes. Principal Component Analysis (PCA) revealed that the first two axes explained 84.27% and 4.89% of the observed variance, respectively (Figure 6). This analysis enabled the identification of three distinct groups of closely linked genotypes in the bi-dimensional space.

The largest group (labelled 1), situated in the bottom left corner of the PCA space, comprised 402 identical samples. The second-largest group (labelled 2), located in the bottom right corner, included 17 samples primarily from the Italian Peninsula, exhibiting columnar habits and adhering to species-specific *P. nigra* characteristics. Notably, ten individuals from Italy, Spain, France, and Poland, positioned in the intermediate PCA space along the PC1 axis between groups 1 and 2, represented intermediate forms displaying variable genetic distance from both groups.

The third group (labelled 3), positioned in the upper left corner of the PCA space, consisted of seven samples with high SNP heterozygosity, originating from trees in a single location in Norway. Phenotypically distinguishable, these trees were considered probable interspecific hybrids. Group 1 specimens were identified as the most prevalent genotype across Europe,

representing the common notion of columnar *P. nigra*. Consequently, a male-flowered tree from group 1, specifically tree 9-39 collected in a parking lot site in Norway, was selected as the reference individual for constructing a genome assembly representative of the taxonomic entity known as *P. nigra* var. 'Italica'.

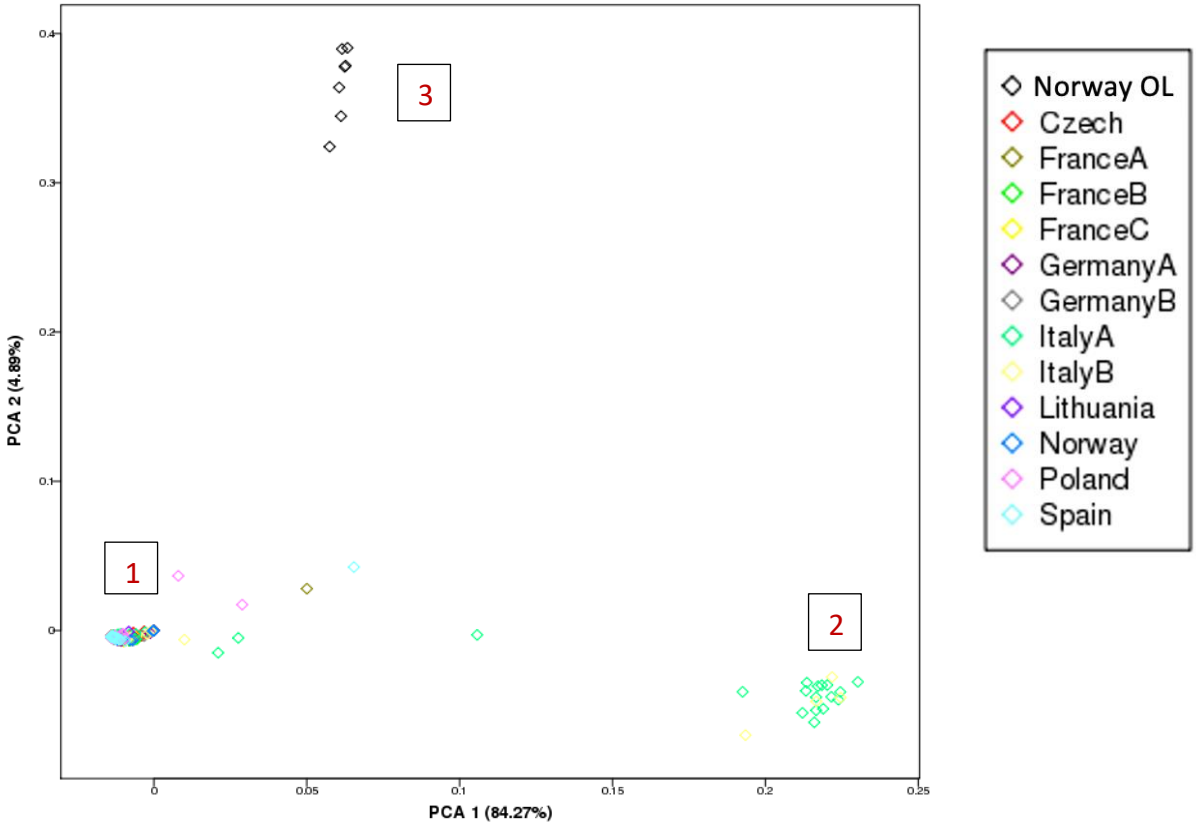


Figure 6: Principal Component Analysis, An examination of 436 columnar *P. nigra* from the EpiDiverse collection. The three primary groups mentioned throughout the thesis are identified by numbers 1,2 and 3. Appendix III lists three subgroups: 1. pure Lombardy poplar, 2. Italian *P. nigra* italica populations, and 3. hybrids of italica with other poplar species. The two red *P. nigra* samples correspond to previously identified *Populus nigra*, non-italica samples. The figure interprets that among EpiDiverse collection, there were three clonal groups. Two of which were true columnar *P. nigra.* Third group was further away from nigra clusters and was suggested to be a interspecies hybrid. The PCA unveiled the most common genetic group as true Lombardy Poplar.

6.1.2 Metrics of the genome assembly

In this study, a multi-platform approach was employed to generate a comprehensive genomic dataset for *Populus nigra* var. italica 9-39. A total of 223 million short linked-reads, obtained through Illumina sequencing (10X Genomics), yielded an estimated coverage of approximately 150X. Additionally, 1.98 million long Oxford Nanopore Technology (ONT) reads, with an N50 length of 29 Kb, provided an estimated coverage of 30X. Furthermore, 100 million Hi-C reads were generated, corresponding to an estimated coverage of 60X. Kmer analysis of linked reads enabled the estimation of a genome size of 425.95 Mb for the individual under investigation.

The linked reads were subjected to assembly using Supernova, resulting in a haplotype-phased assembly spanning 422.52 Mb. This assembly comprised 17,463 contigs with an N50 of 2.9 Mb, and the largest contig measured 13.6 Mb in length. Notably, 77 contigs surpassed the 1 Mb threshold. In the Supernova assembly, 10-bp gaps were introduced at the intersections of contiguous sequences of the same haplotype.

Simultaneously, long-reads were assembled using WTDBG2 software and subsequently polished with Illumina reads, resulting in an assembly of 390.47 Mb. This assembly comprised 1,326 contigs with an N50 of 1.6 Mb, and the largest contig extended to 15.9 Mb. Ninety-four contigs exceeded 1 Mb, indicating a robust long-read assembly strategy. However, the nanopore sequencing approach introduced larger gaps (greater than 500 bp) due to missing sequence data of known length.

Integration of information from both assemblies culminated in a hybrid assembly, encompassing a total length of 417 Mb. This hybrid assembly consisted of 9,677 scaffolds with an L50 of 20 and an N50 of 6.3 Mb, as detailed in Table 3. This multi-platform genomic approach facilitates a more comprehensive understanding of the structural and functional aspects of the *Populus nigra* var. italica 9-39 genome.

Table 3: Basic statistics for individual assembly method i.e. 10X and ONT sequence assembly. V1 assembly denotes the error-corrected, combined version of the two.

Parameter	**10X only assembly**	**ONT Assembly**	**Merged V1 assembly**
Assembly size	422.52 Mbp	390.47Mb	417 Mbp
Total no. of scaffolds	17463	1326	9677
Scaffold N50	2.9Mbp	1.58Mbp	6.34 Mbp
Longest scaffold	13.6Mbp	15.85Mbp	17.68 Mbp
No. of scaffolds > 1M nt	77	94	60

Finally, Hi-C-based scaffolding using SALSA scaffolder V2.2 (Ghurye et al., 2019) generated 9,533 scaffolds with L50 of 13 and N50 of 9.5 Mb. The 500bp gaps in the scaffolded assembly represent the contig scaffolding junction. In the final assembly (denoted as V2 assembly, Table 4), gaps of 100bp stretch exist that represent missing data of unknown length.

Table 4: Scaffolding of V1 assembly using HiC reads, manual adjustments, and organelle integration resulted in the final integrated assembly V2. The stats are described at both contig and scaffold level.

Sample	*P. nigra var Italica*	
Assembler	Merged assembly with Hi-C Scaffolding	
Level	**contigs**	**scaffolds**
Total (bp)	406246012	417754133
Mean Size (bp)	37755	43662
Max (bp)	4339512	33746648
Min (bp)	49	1000
Sequences (#)	10760	9568
N50 (#)	145	13
N50 (bp/L50)	695456	9493943

Synteny analysis was conducted to establish a comparative relationship between the *Populus nigra* var. italica 9-39 assembly and the assembly of its closely related species, *Populus trichocarpa* (Figure 7). The analysis revealed a high degree of synteny, indicating that our scaffolds align seamlessly with those of the sister species. Remarkably, the scaffolds effectively reconstructed contiguous sections of the original chromosomes, providing comprehensive coverage across nearly all genomic regions. The examination also identified two misassembled scaffolds, specifically scaffold 4 and scaffold 8, requiring corrective measures due to their erroneous orientation. Additionally, a 193,269-bp long chloroplast sequence and a 520,494-bp long mitochondrial sequence were successfully identified, contributing valuable insights into the organelle genomic content.

To gauge the completeness of our assembly, we utilized statistics derived from the conserved orthologs gene set (BUSCO; Manni et al., 2021). The analysis indicated that 82% of genes were present in a single copy and were considered complete. Another 13.3% were also complete but were identified in duplicate. Merely 1% of the BUSCO genes exhibited partial reconstruction or fragmentation, while 3.5% were entirely absent from the assembled sequence (Table 5).

In addition, a K-mer spectrum analysis was employed to assess the assembly's coherence with the content within the reads used for its creation. The K-mer "assembly spectra copy number plot" distinctly illustrated two peaks, revealing heterozygous and homozygous content distributions within our assembly (Figure 8). This detailed genomic scrutiny provides crucial insights into the structural integrity and completeness of the *Populus nigra* var. italica 9-39 assembly, laying the foundation for subsequent functional analyses.

Table 5: BUSCO approach to estimate genome completion by looking for analogous genes in V2 assembly. The stats describe the percentage of complete & single-copy genes, complete but duplicated genes, fragmented genes and missing genes.

Description	Value
Assembler	Merged assembly with Hi-C Scaffolding
Complete and single-copy BUSCOs (S)	82.20%
Complete and duplicated BUSCOs (D)	13.30%
Fragmented BUSCOs (F)	1%
Missing BUSCOs (M)	3.50%

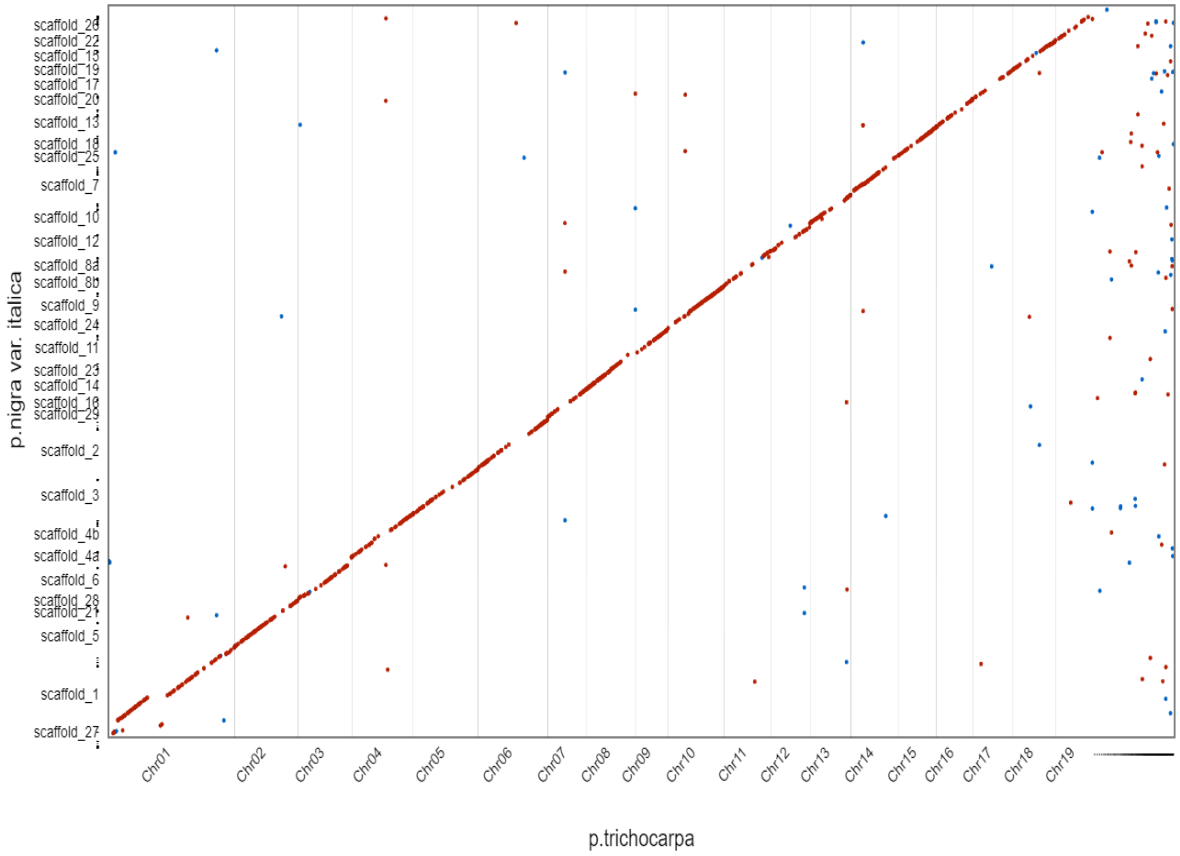


Figure 7: Alignment Graph depicting the contiguity of our V2 assembly's scaffolds in comparison to the 19 chromosome of related species Populus *trichocarpa*. The scaffolds in our *P. nigra var. italica* assembly are depicted in the Y axis and match to the 19 chromosomes present in the *P. trichocarpa* v3 assembly on the X axis. The graph suggests that our assembly was in contiguity with the *trichocarpa* genome, and the fragments can be easily converted into the chromosome scale assembly using a guided genome.

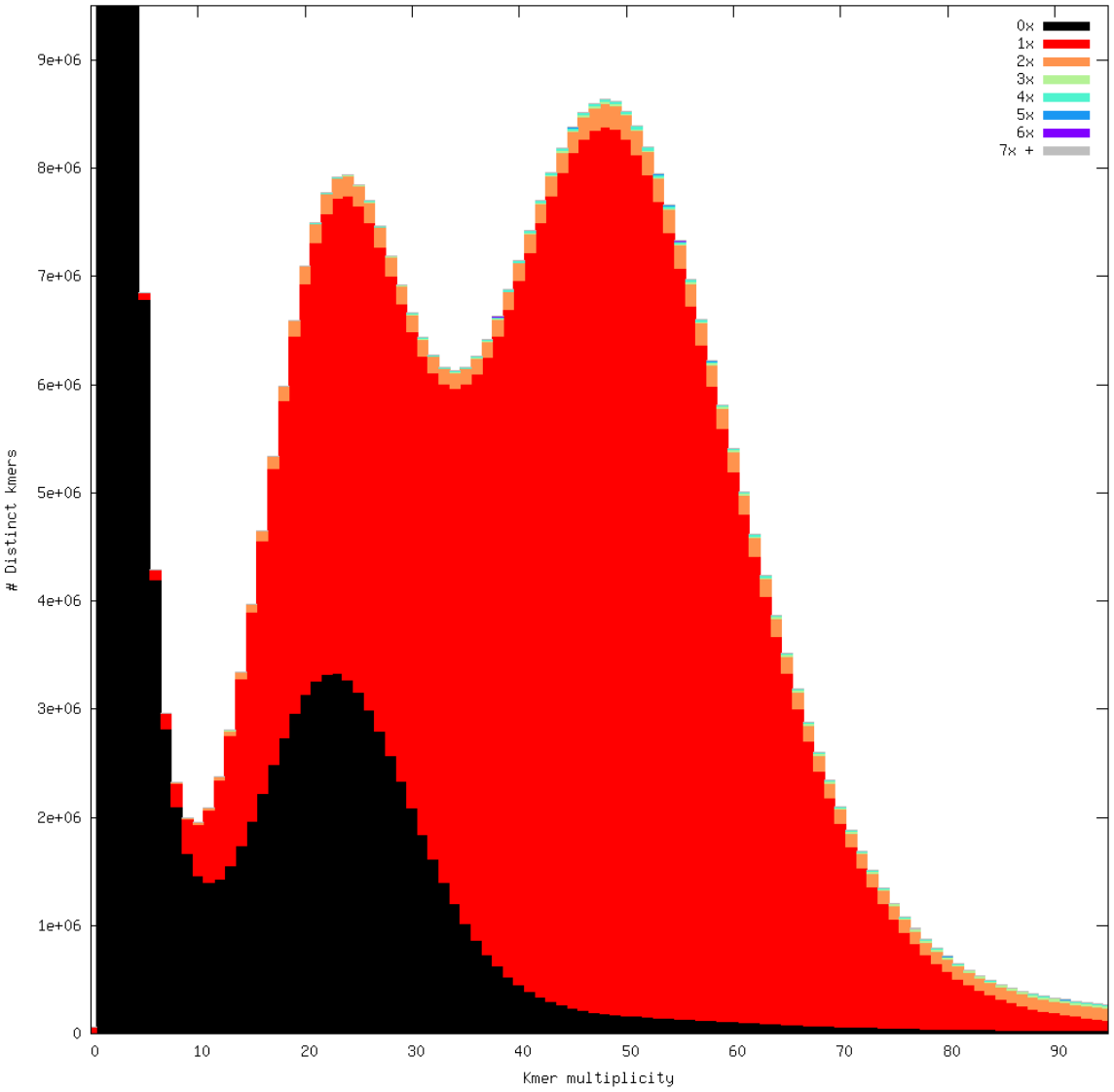


Figure 8: Raw illumina vs. final assembly K-Mer analysis plot. The Y axis reflects the number of distinct k-mers in the alignment, while the X axis denotes kmer multiplicity. The fist peak at around 25X shows the heterozygous content and the peek at 50X shows the homozygous content concluding that the two alleles are equal in size. The black peek shows the lost content i.e. the heterozygous content that was lost when the diploid bubble collapsed in assembly.

6.1.3 Transposable element identification

In order to unveil the repeat elements within the genome of *Populus nigra* var. italica, a meticulous approach combining de novo and homology-based predictions was employed, with subsequent integration for a comprehensive assessment. De novo transposable element (TE) identification utilizing EDTA led to the masking of 310,890 genomic intervals, encompassing 80.07 Mbp of DNA transposable elements, 71.17 Mbp of Long Terminal Repeat (LTR) retroelements, and 1.93 Mbp of other elements. Collectively, these predictions covered 36.72% of the genome. However, the fragmented nature of these predictions indicated a certain degree of imprecision in characterization, primarily attributed to the absence of complete TE features such as long terminal repeats and inverted repeats.

An alternative approach involved leveraging the RepBase database to map known TEs onto our genome, employing more stringent criteria. This method, grounded in a reference library inclusive of TEs from sister species like *P. trichocarpa* and *P. deltoides*, yielded the identification of 231,944 intervals, comprising approximately 62 Mbp of transposable elements, covering 29.9% of the genome. While highly reliable, this library exhibited some fragmentation and incompleteness, with certain TEs absent compared to the initial de novo prediction.

Furthermore, the de novo method Repeatmodeler was employed, resulting in TE predictions covering approximately 43.81% of the genome. These predictions proved to be the most comprehensive, offering accurate de novo predictions and family classification in comparison to predictions from sister species.

The last reference-based library, obtained from a collaborative effort with the University of Perpignan, France, utilized an in-house INRA TE database for TE prediction in our genome. Regrettably, due to its unpublished nature, specific details regarding this tool cannot be disclosed. A detailed presentation of the statistics for each library is provided in Table 6 and Figure 9, facilitating a comprehensive understanding of the diverse approaches employed in identifying repeat elements within the *Populus nigra* var. italica genome.

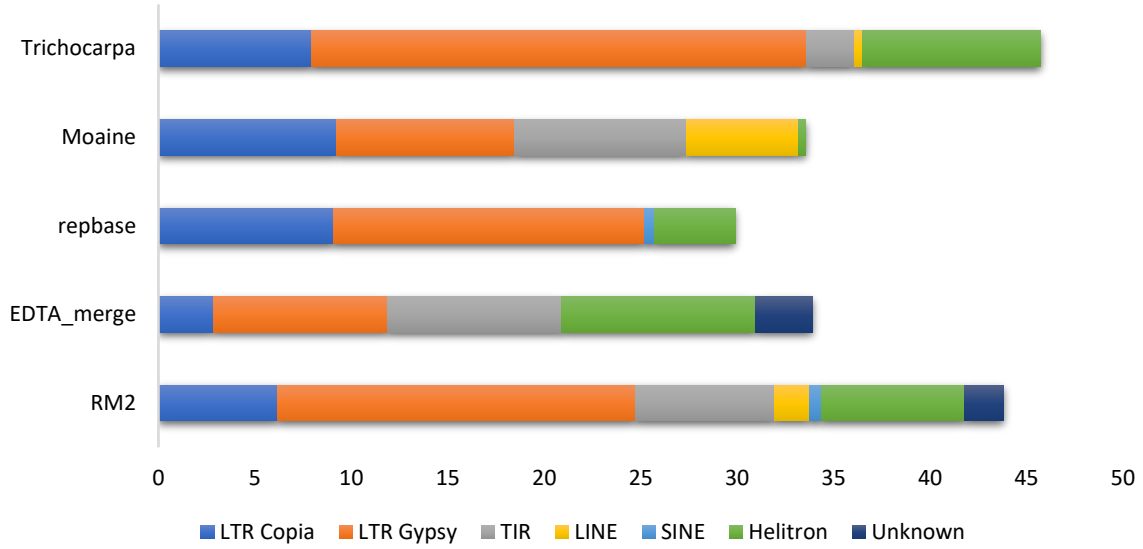


Figure 9: The graph shows the TE family estimates from each approach before integrating. Each major TE is represented on the X axis. In the Y axis, families are categorized using several methods. This comparison compares prediction completeness and accuracy using Repeatmodeler2 (RM2), EDTA-based library (EDTA_merge), repbase (Repbase DB. Based library), Moaine (RexDB based library), and trichocarpa (comparison of the identical TE families contained in *P. trichocarpa* genome V3). The graph interprets that RM2 predictions were most complete but repbase predictions more closely represented family classifications to trichocarpa genome.

Table 6: TE family distributions in repeat libraries developed using various methodologies. This fraction shows the percentage of the plant's genome covered by each method to the entire genome. The table is true numeric representation of the graph above.

% Covered	**RM2**	**EDTA_merge**	**repbase**	**Rexdb**	***Trichocarpa***
LTR Copia	6.17	2.84	9.05	18.48 (total	7.94
LTR Gypsy	18.55	9.05	16.15	LTR)	25.68
TIR	7.22	9.03	0	8.87	2.46
LINE	1.79	0	0.016	5.84	0.41
SINE	0.66	0	0.49	0	0
Helitron	7.42	10.03	4.2	0.36	9.24

Unknown	2	2.97	0	0	0
Total	43.81	33.92	29.906	33.55	45.73

The culmination of our efforts involved the synthesis of the final TE library, achieved through the amalgamation of repbase-based and EDTA de-novo libraries, coupled with Rexdb-based classification for precise TE descriptions. TEs absent in the reference-based predictions underwent cross-validation with Repeatmodeler and another reference-based library to ascertain their accuracy as true positives. Subsequently, characterization was executed, and the detailed outcomes are presented in Table 6.

The resulting TE library provided coverage for approximately 34.76% of the genome, with the largest subset attributed to retrotransposons covering about 13.17% of the total genome. Noteworthy findings indicated that Class II elements exhibited greater fragmentation compared to Class I elements within our library. Additionally, 1.54% of the genome remained ambiguously classified (Table 7; Figure 10). This comprehensive TE library not only enriches our understanding of the repetitive landscape within *Populus nigra* var. italica but also serves as a valuable resource for future genomic analyses.

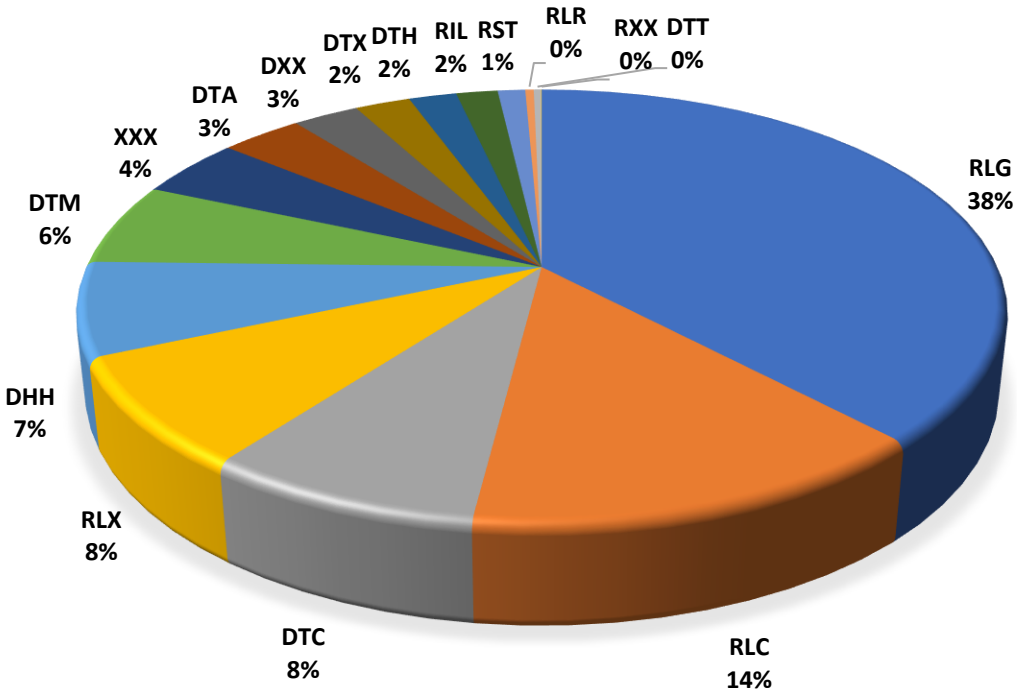


Figure 10: Classification of elements in the final TE library based on our predictions. The fraction of each superfamily is represented in comparison to total no. of repeats. Class I LTR Gypsy elements (RLG) accounted for the greatest proportion of projected TE, followed by Class I LTR Copia (RLC), Unknown Class I LTR (RLX), Class II DNA Helitron, and so on. (Please read Acronyms for a detailed description of each three-letter code)

Table 7: Abundance of different superfamilies in the final TE library created utilizing the library merging approach. The table shows the number of base pairs covered by each class in the genome as well as the number of fragments in which each class was predicted. The table highlights that the DNA elements were more fragmented than retrotransposons.

Class	Family	Base pairs covered in reference	No. of fragments
Class I elements	RLG	55056938	39840
Retrotransposons	RLC	20459445	17628
	RLX	11813581	22030

Total	RIL	2520441	1634
92069144bp	RST	1665815	4278
	RLR	511204	302
	RXX	41720	118
Class II	DTC	12044444	14026
DNA elements	DHH	10117671	17220
Total	DTM	8882821	18012
46696474bp	DTA	4946615	9734
	DXX	4013020	4528
	DTX	3358237	5796
	DTH	2894913	7766
	DTT	438753	1422
Unknown	XXX	6438514	15449

The conclusive Transposable Element (TE) library projections exhibit a congruence with the TE composition observed in other poplar species, falling within the range of 32% to 45% of the genome. Notably, the largest contingent within the TEs is represented by LTR gypsy-type retrotransposons, constituting 38% of the total repeat element content. This prevalence aligns closely with the frequencies observed in other poplar genomes, providing a comparative perspective on the repetitive landscape shared among these species. The meticulous compilation of this TE library enhances our understanding of the genomic dynamics within *Populus nigra* var. italica and contributes to the broader comprehension of TE compositions across the *Populus* genus.

6.1.4 Gene Annotation

We obtained 40,988 gene models, consisting of a total of 212,972 exons. Additionally, we predicted 9,250 5’ UTRs and 6,469 3’ UTRs (Table 8). The gene predictions for our genome exhibited comparability with those of the *P. trichocarpa* genome in terms of completeness, covering 31.15% of the genome with genetic regions (both introns and exons), while the

trichocarpa genome covered approximately 29.58%. However, the number was lower than the predicted gene count for *the P. trichocarpa* genome, estimated at 41,335 genes (Table 9).

Functional annotation was conducted using the automated tool BLAST2GO. Through this method, approximately 98% of the predicted gene models showed similarity (E-value < 1e-8) to the non-redundant (NR) set of plant proteins from NCBI. Nearly 71.5% of the predicted genes received GO annotation, and 91% of the annotations demonstrated similarity to other sister species, such as *Populus trichocarpa* and *Populus deltoides*.

We annotated a total of 34,205 proteins from *P. nigra*. This count is comparable to that of *P. trichocarpa* (34,699) and *P. euphratica* (36,426), but significantly lower than *P. deltoides* (44,853) and *P. alba* (40,213). Based on these comparisons, we conclude that our sequencing and assembly methods were successful in producing a genome and annotations (Li et al., 2022)

Table 8: Stats of the gene annotations generated using the MAKER-P pipeline. The table highlights gene models as well as each gene features like UTRs, exons as well as functional annotations.

Features	**Counts**
Gene models	40,988
exons	212,972
5’ UTR	9,250
3’ UTR	6,469 (3,424 both UTRs)
Genes with functional description	40,377
Genes which are functionally described based on *Populus sp*.	37,298
Genes with predicted protein family	34,205
Genes with ontology terms	29,306

Table 9: Comparing *P. nigra var. Italica* and *P. trichocarpa* Gene Models with Our Predicted Gene Models: Our analysis reveals a slightly lower number of genes in our predictions compared to *trichocarpa* genes. However, our predicted gene models exhibit higher average gene length, consistent with total gene content in the genome, including exon length.

	Nigra Stats	***Trichocarpa* Stats**
No of genes	40988	41335
Average gene length	3175.5	3107.4
Median gene length	2032.5	2427
No. of exons	212972	429455
Average exon length	219.8	210.4
Median exon length	131	124

6.2 Population genomics analysis

6.2.1 Genomics-based definition of the boundaries of the taxa *P. nigra var. italica*

To gain deeper insights into the extent of genetic diversity within columnar *P. nigra* and its relation to global genetic diversity within the *P. nigra* species and the broader *Populus* genus, we initially compared our sets of columnar *P. nigra* (EpiDiverse and CREA-FL sets) with various *Populus* species (26 species from NCBI). Subsequently, we compared these sets with wild populations of *P. nigra* (WatBio collection of European natural populations).

We integrated informative sites from SPET sequencing data that were covered in all three sample sets of *P. nigra* (columnar *P. nigra* in the EpiDiverse and CREA-FL collections, as well as wild *P. nigra* trees in the WatBio collection) with overlapping informative sites in a set of 26 resequenced *Populus* species available at NCBI. A total of 15,280 SNPs (MAF >0.05) were identified in this set of genotypes.

The first two axes of a Principal Component Analysis (PCA) explained 24% and 10.2%, respectively, of the observed variation (Figure 11). This analysis revealed that all columnar *P. nigra* in the EpiDiverse and CREA-FL collections, except for seven trees, form a single

contiguous cluster of individuals without any apparent discontinuity on the PCA space with the surveyed natural diversity of *P. nigra* and is distinct from all other trees in the population.

All columnar *P. nigra* from the EpiDiverse collection belonging to Group 1 and Group 2 in Figure 6 remain within the range of variation of the *P. nigra* species (Figure 11). Genomic data support their botanical classification in the taxa *P. nigra*. However, the seven columnar *P. nigra* from the EpiDiverse collection belonging to Group 3 fall short of the range of variation observed in the *P. nigra* species. Their position on the PCA space may suggest they are intersectional hybrids between *P. nigra* and one or more other *P. nigra* species of the section Tacamahaca (as observed in the generation of *Populus* × *sibirica*) or between *P. nigra* and *Populus* × *canadensis* (a hybrid between *P. nigra* and *P. deltoides*).

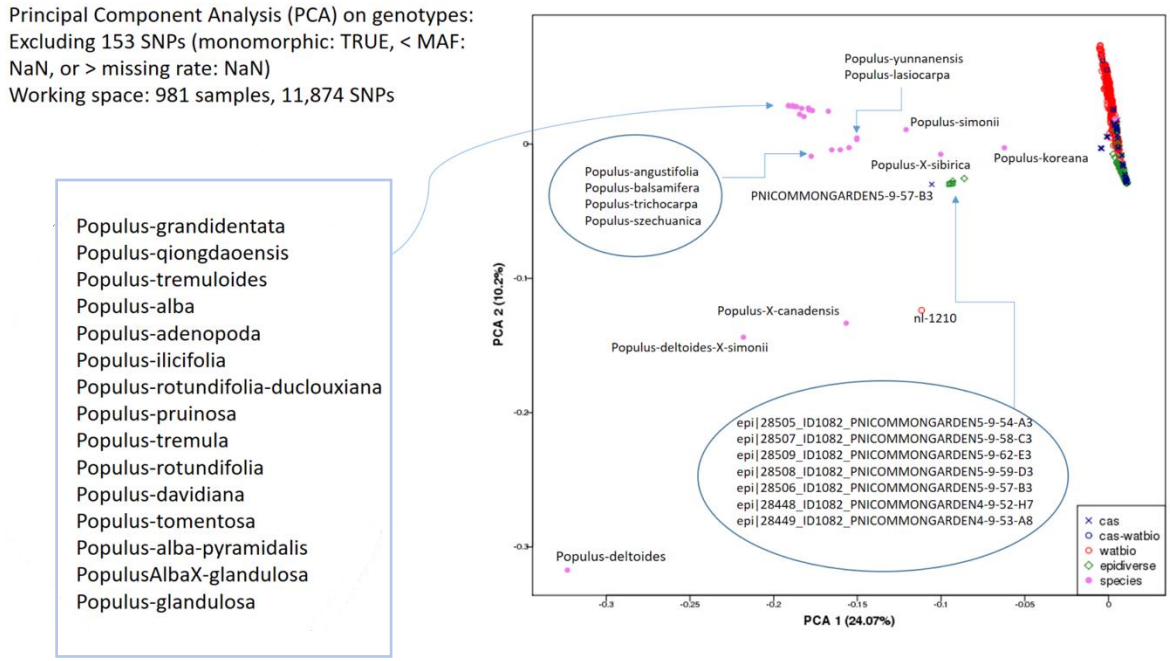


Figure 11: Principal Component Analysis of columnar *P. nigra*, represented by green diamonds (EpiDiverse collection) and blue crosses (CREA-FL collection); spontaneous *P. nigra* trees, represented by red circles (WatBio collection) and other *Populus* species, represented by pink dots, genotyped at 15280 variant sites. X axis represent Eigen vector 1 representing 24.07% variance and Y axis explains Eigen vector 2 with 10.2% variance.

6.2.2 The columnar phenotype in *P. nigra* has at least two independent origins

We combined sequencing data from three sample sets for targeted genotyping, focusing specifically on 5,000 overlapping loci positions. Two sets exclusively comprised columnar *P. nigra*, represented by the EpiDiverse collection (excluding interspecific hybrids), and the CREA-FL collection. The third set, represented by the WatBio collection, included exclusively wild *P. nigra* trees from natural sites across Europe (Scaglione et al., 2019).

A total of 6,416 SNPs (MAF >0.05) were identified in this set of genotypes. The first two axes of a Principal Component Analysis (PCA) explained 33.4% and 7.4%, respectively, of the observed variation (Figure 12). This analysis enabled us to delineate, on the bi-dimensional space, the relationships between two clonal groups of male and female black columnar poplars (indicated as Group 1 and Group 2 in Figure 6) and European natural populations of *P. nigra*. Both clonal groups lay outside the PCA space defined by the observed variation in the European natural populations of *P. nigra* surveyed by WatBio. Their position on the PCA space suggests that the two clonal groups are unrelated to one another. However, groups of black columnar poplars lay only slightly apart from the PCA space populated by the European natural populations of *P. nigra*, indicating they likely originated from related populations of the same species.

Using a threshold of 0.01 genetic distance (genD) to define clonal trees derived by vegetative propagation from a single genotype, we estimated that 463 *P. nigra* columnar trees in the PCA of Figure 12-A belong to either of the clonal groups. The typical male genotype forming one of the clonal groups (Group 1 in Figure 6) is conserved in the CREA-FL collection under the commercial designation of clone ‘San Giorgio’. It likely corresponds to the specimen more narrowly associated worldwide with the common notion of Lombardy poplar and thus with *P. nigra* var. italica. The genotypes of columnar *P. nigra* trees in the EpiDiverse and CREA-FL collections that are not clonal replicates of either ‘San Giorgio’ or the female clonal group are positioned in the PCA at intermediate positions, either between ‘San Giorgio’ and the female clonal group or between either clonal group or natural populations of European wild *P. nigra*.

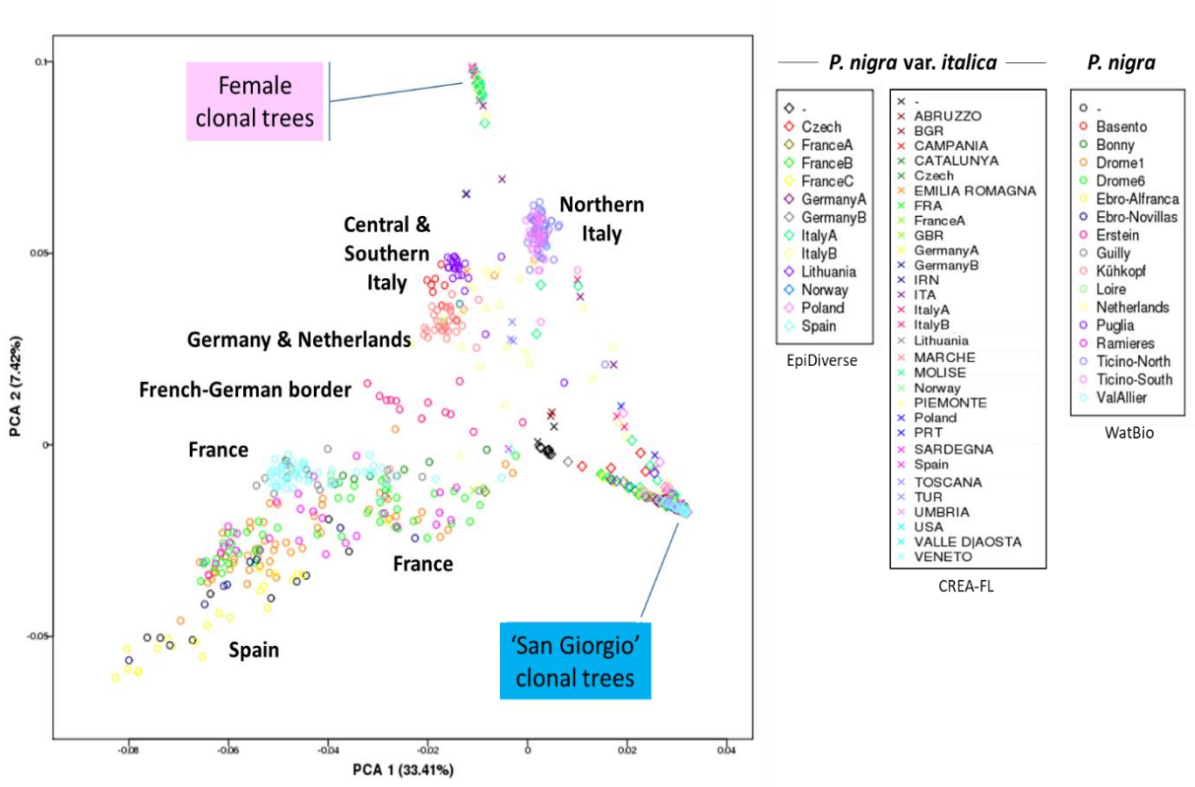


Figure 12-A: Principal Component Analysis of 953 genotyped *P. nigra* trees with 6,416 SNP variants. Circles represent 434 spontaneous *P. nigra* trees from natural populations in Southern and Central Europe (WatBio set); diamonds represent 429 *P. nigra var. italica* trees from the EpiDiverse collection, excluding 7 interspecific hybrids (group 3 in Figure 6), and crosses represent 60 *P. nigra var. italica* trees from the CREA-FL collection. The colours represent the nation or location where the specimen was collected. other as well.

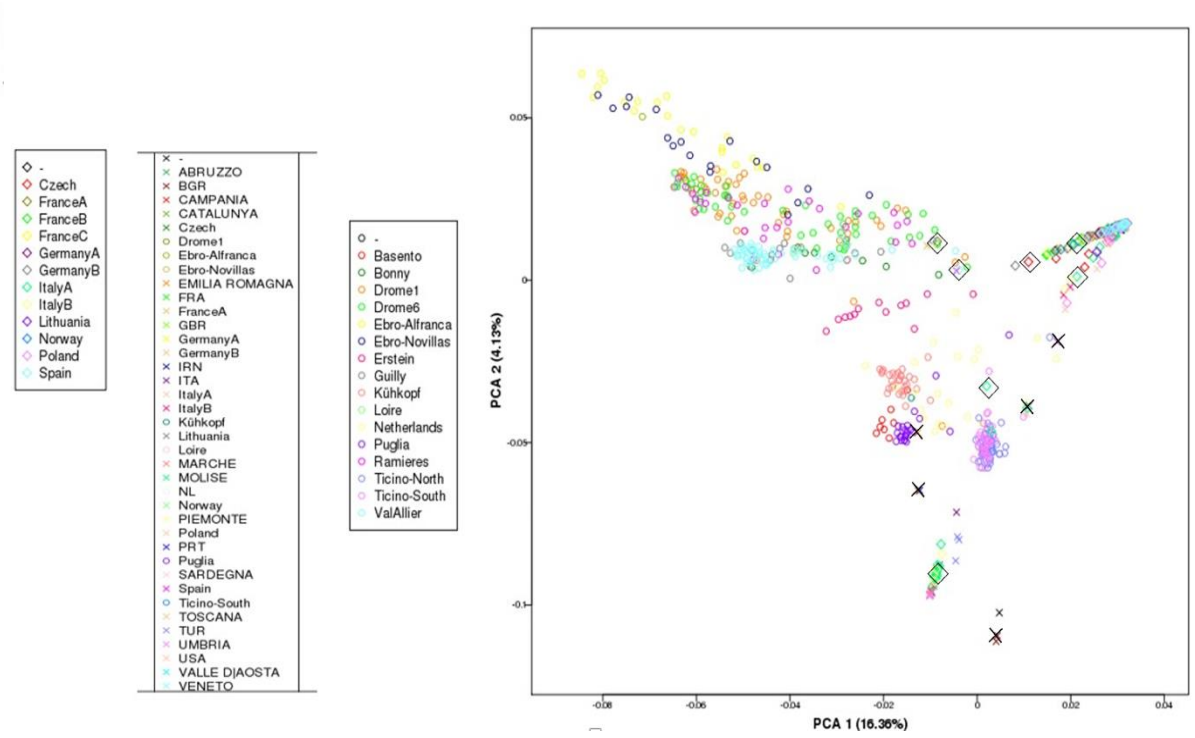


Figure 12-B: Samples selected for resequencing analysis to capture maximum genetic diversity and be able to compare columnar and non-columnar samples. The PCA is an extension of figure 12-A representing the same distribution. The black crosses (samples from CREA-FL collection) and squares (samples from EpiDiverse collection) in this graph represents the samples that were selected fore resequencing analysis based on their clonal distribution. The idea was to take samples from each clonal groups as well as samples that fall in between each group to capture maximum diversity.

6.3 Genealogical relationships in *P. nigra var. italica*

To test all the hypotheses formulated in the previous two paragraphs, we resequenced a total of 18 *P. nigra* trees with approximately 30X coverage. This set included 3 trees from the 'San Giorgio' clonal group (Group 1 in Figure 6), including the accession maintained under the commercial name of 'San Giorgio'. Additionally, one individual from the female clonal group (Group 2 in Figure 6) and 14 distinct genotypes were sequenced. We also included one individual from the putative group of *P. nigra* var. italica vs × *Populus* sp. Hybrids (Group 3 in Figure 6). Four previously sequenced *P. nigra* individuals not classified as pyramidal (Pinosio

et al., 2016) served as negative controls, along with the reference sample. Following alignment, SNP calling using GATK, and quality filtering, we identified about 18.87 million SNPs. These SNPs were employed for haplotype-sharing analysis.

The pseudo-chromosomal plots of pairwise IBD sharing (Appendix III) revealed that trees within either clonal group ('San Giorgio' and the female clonal group) possess identical genomes, confirming their derivation within each group by vegetative propagation. 'San Giorgio' and the female clonal group, while unrelated, share small and fragmented haplotypic blocks, indicating several generations of separation from their most recent common ancestors (Figure 13). These two groups represent independent lineages of columnar *P. nigra*. Other unique columnar *P. nigra* genotypes exhibit close genealogical relationships with the two clonal groups.

The 'San Giorgio' clonal group has produced at least ten offspring, with seven represented by resequenced trees. Notably, 3-04 and 9-57, an interspecific hybrid in Figure 6, exhibit a PO relationship with 'San Giorgio' but do not share the maternal parent, signifying a half-sib relationship. 3-04 is identical or a clonal replicate of a sequenced accession known as 'Blanc de Garonne,' a commercially used clone in plantation forestry in France. N424 and 4-13-23 are full-sibs and share a PO relationship with 'San Giorgio'.

Similarly, N644, N645, and N646, three accessions collected in Bulgaria, are full-sibs and share a PO relationship with 'San Giorgio.' The 2-12 genotype has a genome composition that suggests derivation from two other 'San Giorgio' offspring not captured by our set, sharing a grandparent-grandchild relationship with 'San Giorgio' and exhibiting a highly inbred genome. Another offspring of 'San Giorgio,' different from those mentioned, is needed to explain the genome composition of the trio 'Gazy,' N451, and N452 (all three accessions originating from Turkey). They are full-sibs, in a grandparent-grandchild relationship with 'San Giorgio,' but not half-sibs with the other descendants mentioned. Lastly, tree 5-13-27 shares a PO relationship with both 'San Giorgio' and 4-13-23, likely representing a backcross between 4-13-23 and 'San Giorgio,' as indicated by its high levels of homozygosity in the genome. Two full siblings (N932 and 31-10-29) exhibit a PO relationship with both 'San Giorgio' and the clonal female group, suggesting hybridization between two unrelated pyramidal lineages of *P. nigra*, one being *P. nigra* var. italica. Another accession (N355) likely has a second-degree relationship with the

clonal female group, possibly a grandparent-grandchild relationship, registered in the CREA-FL collection as 'Bordils' (Figure 14). This genealogical data suggests at least two different origins for the columnar growth habit in *P. nigra* var. italica (Figure 14).

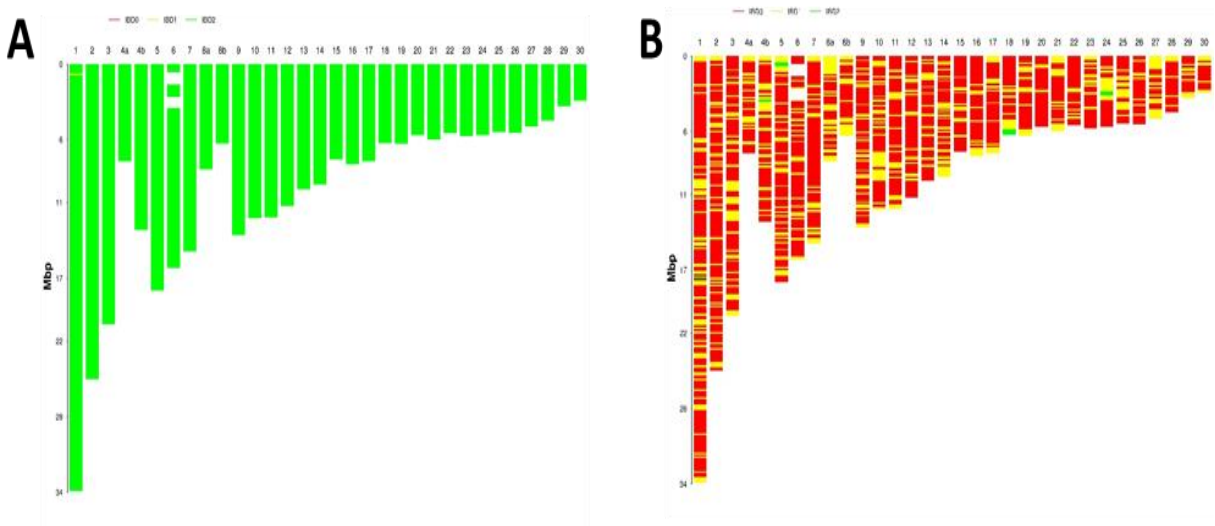


Figure 13: Plot of IBD across 2402 genomic windows over 33 scaffolds of the *P. nigra var. italica* genome between two clonal trees of the 'San Giorgio' clonal group (panel A) and one tree of the 'San Giorgio' clonal group and one of the female clonal groups (panel B). Green genomic windows denote chromosomal fragments in IBD=2. Yellow genomic windows denote chromosomal fragments in IBD=1. Red genomic windows indicate chromosomal fragments in IBD=0. X axis defines the scaffold coordinates in Mbp, scaffolds are presented in increasing order along the Y axis according to our V2 assembly.

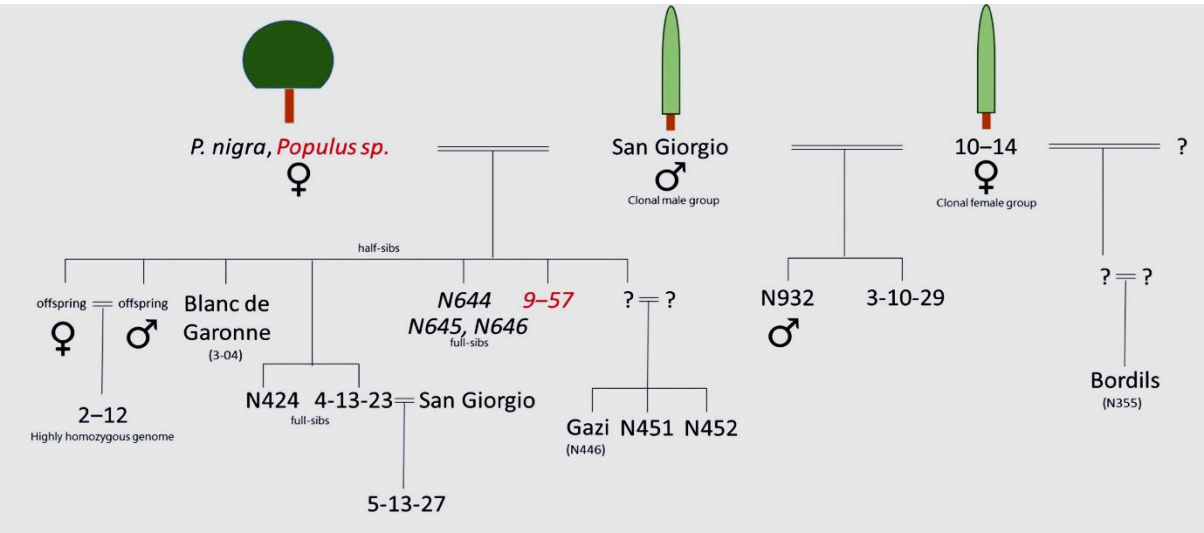


Figure 14: Graphical representation of the pedigree relationship of the resequenced samples. The graphic depicts two primary groups of male and female columnar poplar samples, as well as the progenies originating from them, both individually and in relation to one another.

6.4 Two independent mutations causing premature stop codons in the TAC gene are associated with the columnar phenotype

To ascertain whether the inferred two independent origins of the columnar phenotype in *P. nigra* var. italica correspond to distinct mutations, we employed a candidate gene approach to identify the causal mutation for the vertical branch angle. A literature search revealed three genes known to be associated with columnar habits in other plant species: the Co gene (Columnar gene) responsible for the columnar phenotype in apples (Otto et al., 2013), the PpTAC1 gene (Tiller Angle Control 1) with higher expression in non-columnar types in *Prunus persica* (Dardick et al., 2013), and the LAZY1 gene, which induces asymmetrical auxin distribution leading to parallel branching in Arabidopsis (Table 10). Recent evidence confirmed the association of the Ortholog of the TAC1 gene with the columnar effect in *P. × Canescens* through CRISPR/Cas9-mediated gene knockout (Fladung, 2021). To identify the orthologs in *P. nigra*, we conducted a tBLASTn search across our genome using the protein sequences from

these genes. The putative orthologs in *P. nigra* are located on three different scaffolds of the *P. nigra* var. italica genome assembly (Table 10).

Published evidence indicated that the mutation is homozygous for the reference individual (Zsuffa, 1974). Gene expression profile differences were used to corroborate these findings. Given the favorable circumstances of dioecy, where poplar genomes are highly heterozygous genome-wide, and the reported homozygosity for the causal factor of the columnar phenotype in the two clonal groups of *P. nigra* var. italica (Zsuffa, 1974), we searched for runs of homozygosity in otherwise heterozygous genomes. Large chromosomal homozygous tracts were found around the genomic region encompassing the TAC1 gene in both 'San Giorgio' and the clonal female group. This led us to investigate allelic variation in the TAC1 gene between the two columnar *P. nigra* genotypes and wild-type *P. nigra*. We identified a stop-gain homozygous A to T mutation at position 7,350,817 of scaffold 7 in the exon 3 of PnTAC1 for 'San Giorgio' (Figure 16, A), resulting in a premature termination codon. The clonal female group carried a homozygous frame-shift single-base deletion at position 7,350,365 of scaffold 7 in the exon 2 of PnTC1 (Figure 16, C), leading to a premature termination codon at position 7,350,691 in exon 3 (Figure 15).

Evidence from RNA-Seq alignments indicated that TAC1 is transcribed in wild-type *P. nigra* accessions 71077 and BEN3 (Bio Project number PRJNA303130, Faivre-Rampant et al., 2016). However, we did not find evidence of the presence of mature mRNA transcripts from read alignments of 'San Giorgio' RNA-Seq libraries obtained within the EpiDiverse project (Figure 16, B). We hypothesize that the truncated transcripts in 'San Giorgio' induce nonsense-mediated RNA decay (Rayson et al., 2012). We further postulate that NMD likely triggers the mechanism resulting in truncated transcripts in the clonal female group as well, although supporting information on transcript abundance is lacking for verification.

Table 10: The table summarises genes associated with the columnar phenotype in other species, the mutation generating this phenotype, their orthologs in our genome, and the BLAST algorithm-based similarity score.

Organism	**Gene Name**	**Region (length Mb)**	**Mutation**	**Orthologues in our genome(bp)**	**Similarity and q-coverage (tblastn)**
Malus domestica	*Co* Gene	Chr 10 18.73–18.92	Ty3/Gypsy @ Chr 10 18.8 Mb	Scaffold_5 15157942-15162890	Q cov: 39.1% E-value: 4e-137 Sim perc: 71.2%
Prunus persica	*TAC1*	Scf 2 23234012-23236974	High expression of the gene	Scaffold_7 7350269-7351530	Q cov: 77.2% E-value: 3e-27 Sim perc: 43.7%
Arabidopsis thaliana	*LAZY1*	Chr 5 4,547,230-4,549,214	Asymmetric auxin distribution	Scaffold_1 11478851-11481767	Q cov: 57% E-value: 2e-31 Sim perc: 42.1%

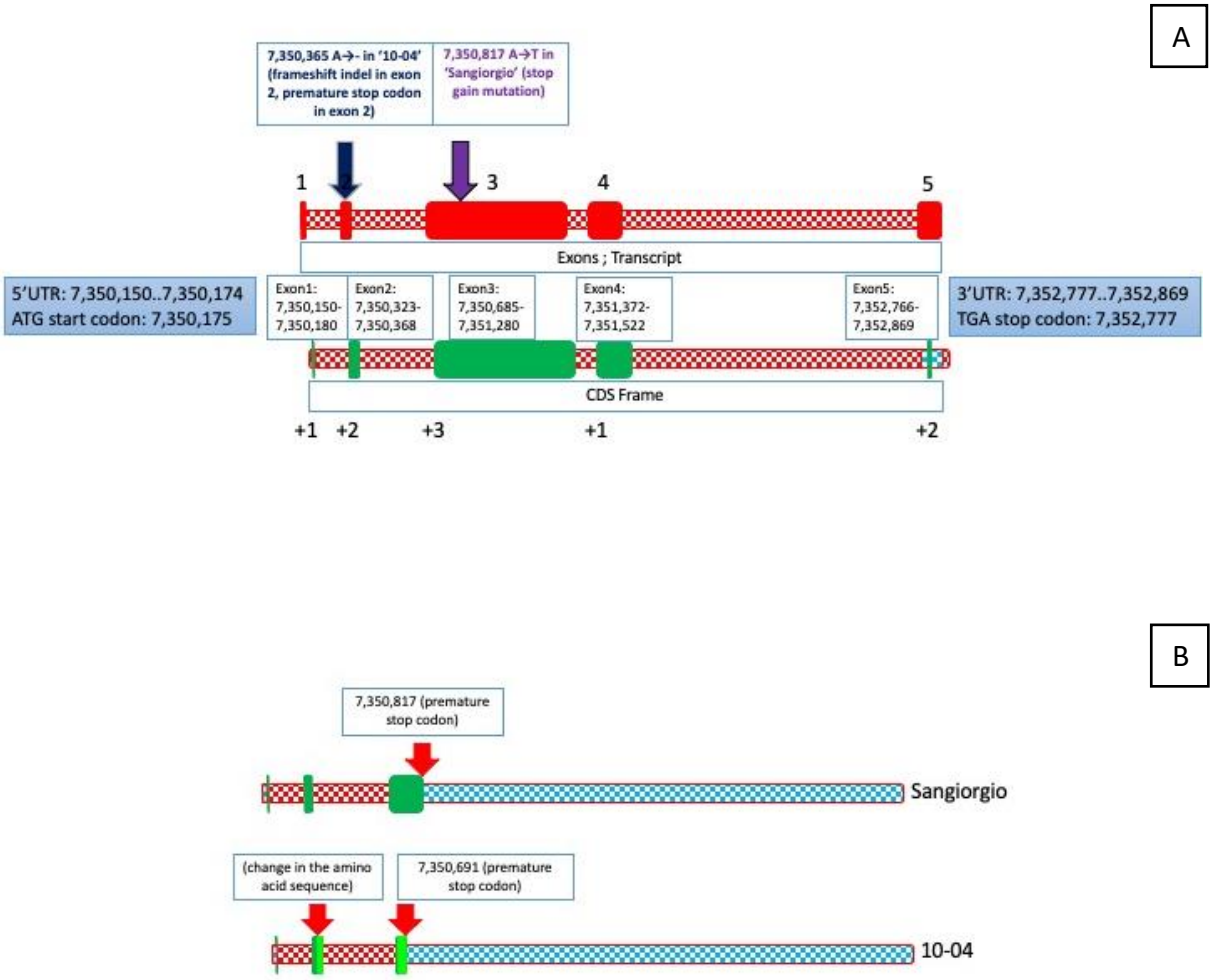


Figure 15: Graphical representation of the structure of PnTAC1 gene. The panel A is the representation of TAC1 gene (above) and protein (below) structure in wild type *P. nigra.* It contains a total of 5 exons. The panel B shows the predicted gene and protein structure in the two groups of columnar genotypes: the San Giorgio sample containing the A to-T non-sense mutation causing premature termination of translation causing loss of function in the gene, and the female clonal group (10-04) containing the frameshift indel causing change in amino acid sequence as well as another premature stop codon in exon 3 which is causing translation termination resulting in the loss of function.

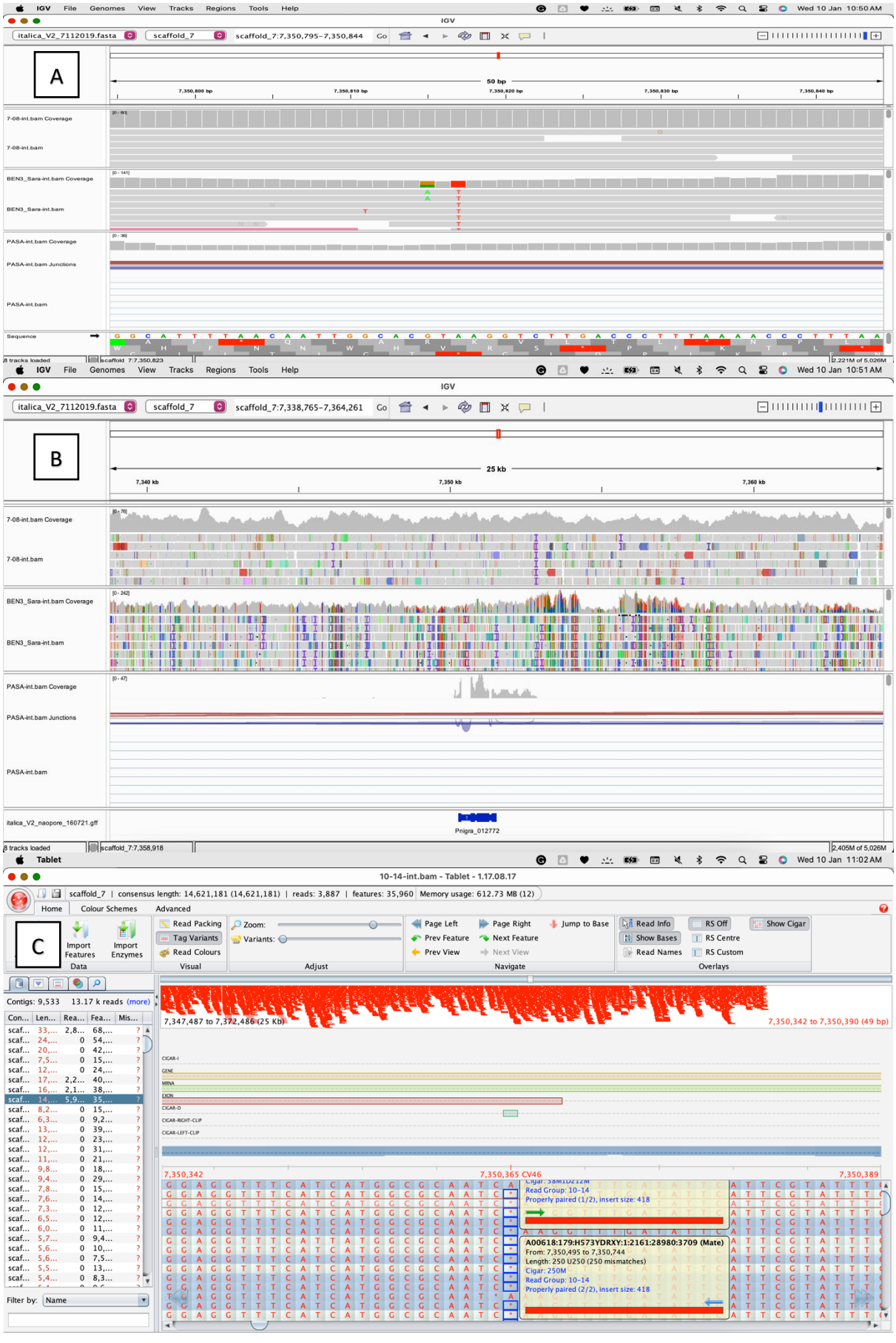

A
B
C
IGV
Tablet
50 bp
25 kb
7-08-int.bam Coverage
7-08-int.bam
BEN3_Sara-int.bam Coverage
BEN3_Sara-int.bam
PASA-int.bam Coverage
PASA-int.bam Junctions
PASA-int.bam
Sequence
italica_V2_naopore_160721.gff
Pnigra_012772
10-14-int.bam - Tablet - 1.17.08.17
Contigs: 9,533
13.17 k reads (more)
7,347,487 to 7,372,486 (25 Kb)
7,350,342 to 7,350,390 (49 bp)
7,350,342
7,350,365 CV46
7,350,389
A00618:179:H573YDRXY:1:2161:28980:3709 (Mate)
From: 7,350,495 to 7,350,744
Length: 250 U250 (250 mismatches)
Cigar: 250M
Read Group: 10-14
Properly paired (2/2), insert size: 418

Figure 16: Image representing the TAC 1 gene orthologue in *P. nigra* var italica specifically focusing on the two variants causing the columnar phenotype. A) Comparing sample 7-08 (San Georgio italica tree) with BEN3 (wild type *P. nigra* from watbio) representing T > A point mutation at position 7,350,817 in sample 7-08, making TAA stop codon. The lower panel is resulting peptide. B) The premature stop codon in exon 3 terminates the transcription in San Giorgio samples resulting in lower mRNA coverage as seen in track 3. C) Represents the female *P. nigra var italica* with frameshift indel at position 7,350,365 causing stop codon downstream.

7 Discussion

7.1 Genomic Complexities and Advancements in Lombardy Poplar: Insights from Assembly and Annotation

The investigation into the Populus genome has unveiled a relatively recent genome duplication event, occurring approximately 60-65 million years ago, affecting nearly 92% of the genome and leading to the divergence of *Populus* from other Salix lineages. Approximately 8000 genes within the trichocarpa genome were identified as paralogous, emphasizing the intricacies arising from this duplication event (Tuskan et al., 2006). The implications of such an event have posed substantial challenges in genome assembly for *Populus* trees due to the introduced redundancy. Employing a hybrid assembly approach, we achieved success in generating a contiguous assembly, establishing collinearity between the *P. trichocarpa* 19 chromosomal molecules and our genome. The 44 largest scaffolds accounted for over 90% of the genome. However, challenges persisted, stemming from unassembled genomic portions resulting from non-coverage in sequencing, the tree's ploidy, and the prevalence of high copy number repeats (Claros et al., 2012). Despite the challenges, the inclusion of high copy number repeats led to the formation of small, challenging-to-assemble scaffolds that effectively minimized gaps.

A comparative analysis with the recently reported poplar genome assembly of *P. euphratica* (Zhang et al., 2020) underscored the notably higher quality of our assembly, offering promising prospects for future research endeavours. The intricacies introduced by genome duplications extended to the realm of genome annotation, evidenced by a lower gene count (40,988) compared to sister species, highlighting the impact of misannotated paralogous genes as single genes. Misassembly significantly contributed to an underestimation of tandem genes, a common issue in plant genomes of draft quality, attributed to limitations in sequencing and evidence (Panchy et al., 2016). Disparities in average gene length hinted at potential variations in genome organization and gene structure (gene coverage in the genome) between *P. nigra var. italica* and *P. trichocarpa*, acknowledging the potential influence of sequencing artefacts and methodological variations.

Despite achieving acceptable quality for future research, continual improvement is imperative for achieving a perfect chromosome-scale phased genome assembly. The incorporation of long-read sequencing holds promise for enhanced haplotype resolution, pseudogene identification, and the delineation of precise boundaries for different gene features. Additionally, employing genome mapping techniques proves beneficial for identifying the exact 19 chromosomes of the clone. Moreover, gene annotations would benefit significantly from additional mRNA-based evidence, enhancing the precision of gene start and end coordinates.

7.2 TE Dynamics in *Populus nigra var. italica*: Unravelling, Resolving, and Propelling Genomic Insight

The identification of Transposable Elements (TEs) within the genome has yielded profound insights into the dynamic intricacies and structural composition of the genome. These elements, actively involved in gene regulation, evolutionary processes, and chromosomal rearrangements, exert a substantial influence on genetic diversity and the adaptability of organisms to environmental changes. The comprehension of the TE landscape holds not only theoretical importance but also practical implications for plant breeding and genetic enhancement initiatives (Razali et al., 2019). However, the inherent limitation of de novo approaches lies in their propensity to produce fragmented predictions, potentially leading to the oversight or mislabelling of certain TEs (Storer et al., 2022).

Our study strategically employed a myriad of techniques encompassing de novo predictions, homology-based mappings using RepBase (Bao et al., 2015), and the integration of separate libraries. This comprehensive approach aimed to yield a more precise and comprehensive TE annotation, addressing the challenges intrinsic to de novo predictions. Within the genome of *Populus nigra var. italica*, our investigation unveiled a substantial presence of TEs, constituting approximately 34.76%. Remarkably, LTR gypsy-type retrotransposons emerged as significant contributors, encompassing about 38% of the overall repeat element content, aligning with observations in other poplar species (Zhao et al., 2022).

To enhance TE annotations within the *P. nigra* var. italica genome assembly, a proposition is made to explore complementary approaches, such as long-read sequencing technologies. This endeavour aims to achieve more comprehensive and contiguous assemblies for repetitive

regions. Envisaging comparative genomic studies with closely related Populus species emerges as a promising avenue, poised to identify conserved and species-specific TE families. The integration of multiple prediction methods, coupled with rigorous experimental validation, holds the potential to foster a more accurate understanding of the repetitive element landscape within the genome, thereby advancing our knowledge of genome evolution in *Populus nigra* and related plant species (Storer et al., 2022). The prevalence of Transposable Elements (TEs) within a substantial fraction of the genome induces extensive genetic variance within populations, consequently fostering genetic diversity (Sahebi et al., 2018; Storer et al., 2022). Genetic diversity, a pivotal mechanism for population adaptation to dynamic environments, enhances the likelihood of survival and reproduction, thereby ensuring population persistence across successive generations (National Biological Information Infrastructure, 2011).

7.3 New findings about the history, origin, and distribution of Lombardy Poplar in Europe

To unravel the genetic relationships within our studied population, Principal Component Analysis was employed. The analysis revealed the presence of multiple clonal groups, with the largest confirmed as the bona fide *Populus nigra* var italica, commonly known as the male Lombardy poplar. The second largest group comprised female columnar *Populus nigra* trees exclusively from Italy, while the third group represented a hybrid columnar poplar. Our results shed light on the widespread distribution and adaptation of columnar *P. nigra* of the italica genotype in the European climate, encompassing both male and female trees.

The research underscored the formidable challenge of identifying true-to-type Poplar trees in this region due to a high number of intra and inter-species hybrids exhibiting similar phenotypes. Our analysis proposes a genetic map for the identification of *Populus nigra* var. Italica among these hybrids.

In the context of climate change, the anticipated alterations in the distribution of major forest tree species in Europe have led to an increased focus on minor native species. This shift in attention aims to comprehend the genetics and ecology of these species for the protection of forest systems (Koch et al., 2022). The Lombardy poplar, identified as *Populus nigra* var. italica,

represents one such species. Previous studies primarily investigated its origin and distribution for selective localized breeding purposes, leaving the genetic origins relatively unexplored (Smulders et al., 2008).

Presumably originating in central Asia, it was introduced to the Lombardy region in Italy, where it propagated as a male fastigiate mutation, subsequently recognized as Lombardy poplar (*Populus Nigra* "Italica" - Plant Finder, n.d.). Our research defined genetic boundaries for *P. nigra* var. italica, offering valuable insights into the genetic diversity and relationships of columnar *P. nigra* from EpiDiverse and CREA-FL collections compared to other *Populus* species (26 species from NCBI) and wild *P. nigra* populations (WatBio Collection) across Europe. Principal Component Analysis (PCA) demonstrated that most columnar *P. nigra* individuals from the EpiDiverse and CREA-FL collections formed a distinct cluster, supporting their classification as *P. nigra* var. italica. Nevertheless, some individuals exhibited differing PCA patterns, hinting at potential intersectional hybrids with other *P. nigra* species or hybrids like *Populus × Canadensis*.

Regarding the geographical distribution of *P. nigra*, our analysis revealed a strong relationship among trees originating from the same region, regardless of clonal classification. Female columnar trees from Italy (north and central) were genetically closer to non-italica clusters from north and central-southern Italy regions than to italica trees from Italy. Conversely, male italica trees formed a distinct cluster based on their columnar phenotype, suggesting that clonal selection for the columnar phenotype is independent of the geographical distribution of *P. nigra* populations, affirming the artificial selection of male italica clones (Smulders et al., 2008).

Furthermore, our study unveiled that the columnar phenotype in *P. nigra* has at least two independent origins represented by two clonal groups of male and female black columnar poplars. These groups are unrelated to each other but genetically linked to European natural *P. nigra* populations, indicating a complex evolutionary history and potential hybridization events contributing to the columnar growth habit.

Our findings support the traditional belief in the single origin of the Lombardy poplar, even though the specific location and date of the event remain elusive (Bean & Clarke, 1989),

presumably occurring in the Lombardy region of Italy in the 17th century. The presence of intersectional hybrids and the relatedness to other Populus species introduces complexity to the taxon's distribution and genetic diversity, with implications for the conservation of *P. nigra* var. italica. This suggests that the species might be less adapted to European conditions than previously assumed (European atlas of forest tree species, n.d.), rendering it more vulnerable to pests, diseases, and climate change.

While our method employed a limited sample size of genotyped individuals, a more extensive sampling across diverse regions is imperative for a comprehensive understanding of *P. nigra* var. italica's phytogeography. Furthermore, additional research is necessary to elucidate the mechanisms underlying the columnar growth habit. Subsequent studies should incorporate genomic data from a broader range of populations and closely related species to unravel the phylogeographical patterns and evolutionary history of columnar *P. nigra.* Functional studies exploring the genetic basis of the columnar phenotype can provide insights into its adaptive significance and ecological relevance. Comparative genomics studies involving other columnar poplar varieties and related species can help uncover whether similar genetic mechanisms underlie columnar growth habits in different lineages. A multidisciplinary approach, integrating genetics, genomics, and molecular biology, is essential to advance our understanding of the genetic basis of the columnar growth habit in Lombardy poplar and related species.

7.4 Genetic Basis of Columnar Phenotype: Insights from TC1 Mutations

Tree architecture is governed by fundamental processes such as primary growth, branching patterns, flowering location, and meristem and shoot mortality (Costes, Lauri, & Regnard, 2006). Among these, branching patterns significantly influence the core structure of the tree, with columnar branch growth offering advantages in high-density planting, minimal pruning, and mechanical harvesting (Petersen & Krost, 2013). This study aimed to unravel genetic differences in the TAC1 gene region, known for its impact on columnar branch angles in other *Populus* species (Fladung, 2021).

The columnar phenotype in *P. nigra var. italic*a was found to be a simple trait controlled by a single gene and mutation events in it. The results revealed the association of at least two independent mutations in the TAC1 gene with the columnar phenotype in black poplar. These mutations, causing premature stop codons, resulted in the downregulation of functional TAC1 protein, crucial for regulating branching angles and contributing to columnar growth. The identified frameshift indel and stop gain mutations altered the polypeptide during translation, leading to the production of non-functional or abnormally truncated proteins (DNA Is Constantly Changing Through the Process of Mutation | Learn Science at Scitable, n.d.).

Further investigation indicated that the premature stop codon triggered transcript degradation through non-sense mediated mRNA decay in the San Giorgio samples. However, for the mutations to impact branching angles, transcript degradation was not essential; the presence of a non-functional protein sufficed. The absence of mRNA data for the female clone raised uncertainty about whether NMD is triggered by this mutation.

Considering TAC1's constitutive expression in plants and its positive correlation with branching angle variation, the inactivation of TAC1 through loss-of-function mutations emerged as a plausible mechanism for the columnar habit in *P. nigra* var. italica (Hollender et al., 2020).

8 Conclusion

This research aimed to understand the genetic basis of the columnar growth habit in *Populus nigra var. italica* and to generate a comprehensive genome assembly for this tree variety. Our population genomics approach revealed that within European populations, the *P. nigra* trees with columnar habits have at least two independent introductions. We determined that the columnar growth habit is not exclusive to clone italica but results from independent mutations in the TAC1 gene, leading to loos of function of the gene and reduced shoot-to-branch angles. This indicates the presence of the columnar phenotype in various hybrids and mutated spontaneous populations of the genus *Populus* (Fladung, 2021). The results provide crucial insights into species dynamics, genetic diversity, and the impact of human breeding, gene flow, and mutations on biodiversity.

The assembly process employed a hybrid approach, successfully overcoming limitations to achieve a highly consistent assembly. The sequencing data and assembly are publicly accessible, contributing to genomic resources. It can be accessed from the following link in ENA and Jbrowse IGA-TS:

https://www.ebi.ac.uk/ena/browser/view/PRJEB44889

https://genomes.igatechnology.com/jbrowse/public/?data=data/populus-nigra-italica/json

Our research expands the realm of genome annotation efforts by focusing on gene models and repeat elements in *Populus nigra var. italica*. Comparative analyses with related species underscored the reliability of the reference genome and annotations, shedding light on genome structure variations across the *Populus* genus.

By uncovering the genetic origins of the columnar growth habit, our work contributes to the broader fields of genomics, evolution, and plant breeding (Tobutt, 1985).

Furthermore, our research focuses on the genetic diversity and origins of the columnar growth habit in *Populus nigra var. italica*. The unique genetic identity and evolutionary history of columnar *P. nigra var. italica* were highlighted, providing insights into its adaptation and

differentiation (Nicotra et al., 2010). Future studies can explore the trait's origins and mechanisms through broader sampling and additional markers, enhancing our comprehension of genetic landscape and evolutionary dynamics. Further deep analysis can also be conducted to explore phenotypic plasticity in different climates and geographical settings.

The research enriches our understanding of forest system evolution and conservation by illuminating the emergence of unique phenotypes in tree populations. This insight informs adaptive management strategies and contributes to species preservation. The methodology employed is a robust framework for similar studies in tree species.

Understanding the genetic basis of the columnar phenotype has significant implications for Lombardy poplar breeding and genetic engineering efforts. Knowledge gained from this study could be used for targeted breeding strategies or genetic modifications. The study's finding that the columnar growth has at least two independent origins indicates the hypothesis of convergent evolution (Maloy & Hughes, 2013) with distinct mutations in the TAC1 gene leading to a similar phenotype in different lineages. Functional studies, such as gene expression analysis and gene knockout (CRISPR-CAS9), would be valuable to confirm the causal relationship between the TAC1 mutations and the columnar growth habit.

9 References

Abril, & Castellano. (2019). Genome Annotation. In Encyclopaedia of Bioinformatics and Computational Biology (pp. 195–209). Academic Press. https://doi.org/10.1016/B978-0-12-809633-8.20226-4

Arima-HiC Kit User Guide for Plant Tissues. (2019, October). Arima-HiC. Retrieved May 30, 2023, from https://arimagenomics.com/wp-content/files/User-Guide-Arima-HiC-for-Plant-Tissues.pdf

Bannoud, F., & Bellini, C. (2021). Adventitious Rooting in *Populus* Species: Update and Perspectives. Frontiers in Plant Science, 12. https://doi.org/10.3389/fpls.2021.668837

Bao, W., Kojima, K. K., & Kohany, O. (2015, June 2). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA, 6(1). https://doi.org/10.1186/s13100-015-0041-9

Bean, W. J., & Clarke, D. L. (1989, March 1). Trees and Shrubs Hardy in the British Isles. John Murray. ISBN 0-7195-2427-X https://doi.org/10.1604/9780719517907

Browning, S. R. (2008). Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes. Genetics, 178(4), 2123–2132. https://doi.org/10.1534/genetics.107.084624

Campbell, M. S., Law, M., Holt, C., Stein, J. C., Moghe, G. D., Hufnagel, D. E., Lei, J., Achawanantakun, R., Jiao, D., Lawrence, C. J., Ware, D., Shiu, S. H., Childs, K. L., Sun, Y., Jiang, N., & Yandell, M. (2013, December 4). MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. Plant Physiology, 164(2), 513–524. https://doi.org/10.1104/pp.113.230144

Carbutt C., 2012. The emerging invasive alien plants of the Drakensberg Alpine Centre, southern Africa. Bothalia, 42(2):71-85. http://www.sanbi.org

Carr, G.W. & Walsh, N.G. (1996) Salicaceae In Walsh, N.G. & Entwisle, T.J. (ed) Flora of Victoria Vol 3. Inkata Press, Melbourne, Victoria.

Claros, M. G., Bautista, R., Guerrero-Fernández, D., Benzerki, H., Seoane, P., & Fernández-Pozo, N. (2012). Why assembling plant genome sequences is so challenging. Biology, 1(2), 439–459. https://doi.org/10.3390/biology1020439

Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., & Kehrer-Sawatzki, H. (2013, July 3). Where genotype is not predictive of phenotype: towards an understanding of the molecular basis reduced penetce in human inherited disease. Human Genetics, 132(10), 1077–1130. https://doi.org/10.1007/s00439-013-1331-2

Costes, E., Lauri, P.É., & Regnard, J.L. (2006). Analyzing fruit tree architecture: Implications for tree management and fruit production. In J. Janick (Ed.), Horticultural Reviews (Vol. 32, pp. 1–61). John Wiley & Sons, Inc.

Dardick, C., Callahan, A., Horn, R., Ruiz, K. B., Zhebentyayeva, T., Hollender, C., Whitaker, M., Abbott, A., & Scorza, R. (2013, June 13). PpeTAC1 promotes the horizontal growth of branches in peach trees and is a member of a functionally conserved gene family found in diverse plants species. The Plant Journal, 75(4), 61plant https://doi.org/10.1111/tpj.12234

DNA Is Constantly Changing through the Process of Mutation | Learn Science at Scitable. (n.d.). https://www.nature.com/scitable/topicpage/dna-is-constantly-changing-through-the-process-6524898/

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19(1), 11-15

EUFORGEN Technical Guidelines for genetic conservation and use for black poplar (Populus nigra). (2003). Bioversity International. https://www.euforgen.org/fileadmin/templates/euforgen.org/upload/Publications/Technical_guidelines/Technical_guidelines_Populus_nigra.pdf

European Commission, Joint Research Centre, Caudullo, G., De Rigo, D., Mauri, A. (2016). European atlas of forest tree species, (G.Caudullo,editor,D.De Rigo,editor,A.Mauri,editor,T.Houston Durrant,editor,J.San-Miguel-Ayanz,edito) Publications Office of the European Union. https://data.europa.eu/doi/10.2760/776635

Faivre-Rampant, P., Zaina, G., Jorge, V., Giacomello, S., Segura, V., Scalabrin, S., Guérin, V., De Paoli, E., Aluome, C., Viger, M., Cattonaro, F., Payne, A., PaulStephenRaj, P., Le Paslier, M. C., Berard, A., Allwright, M. R., Villar, M., Taylor, G., Bastien, C., & Morgante, M. (2016, March 24). New resources for genetic studies *in Populus nigra:* genome-wide SNP discovery and development of a 12k Infinium array. Molecular Ecology Resources, 16(4), 1023–1036. https://doi.org/10.1111/1755-0998.12513

File:*Populus nigra Italica* - Golan H eights.jpg. (2022, January 17). Wikimedia Commons. Retrieved 09:30, August 30, 2023 from https://commons.wikimedia.org/w/index.php?title=File:Populus_nigra_Italica_-_Golan_Heights.jpg&oldid=622203301

Fladung, M. (2021, November 23). Targeted CRISPR/Cas9-Based Knock-Out of the Rice Orthologs TILLER ANGLE CONTROL 1 (TAC1) in Poplar Induces Erect Leaf Habit and Shoot Growth. Forests, 12(12), 1615. https://doi.org/10.3390/f12121615

Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. (2011, January 31). Considering Transposable Element Diversification in De Novo Annotation Approaches. PLoS ONE, 6(1), e16526. https://doi.org/10.1371/journal.pone.0016526

Frison, G. (2014, December). Pioppi neri con portamento fastigiato. Http://Www.Giuseppefrison.It/Wp-Content/Uploads/2016/05/Pioppi-Neri-Con-Portamento-Fastigiato-Ed.-Cpmp.Pdf

Gaudet, M. (2006). Molecular approach to dissect adaptive traits in native European *Populus nigra L.*: Construction of a genetic linkage map based on AFLF, SSR and SNP markers. https://core.ac.uk/download/pdf/41154639.pdf

Genetics and Molecular Biology primer. (n.d.). https://web.archive.org/web/20050411182216/http://members.cox.net/amgough/Fanconi-genetics-genetics-primer.htm

Ghurye, J., Pop, M., Koren, S., Bickhart, D., & Chin, C. S. (2017, July 12). Scaffolding of long read assemblies using long range contact information. BMC Genomics, 18(1). https://doi.org/10.1186/s12864-017-3879-z

Ghurye, J., Rhie, A., Walenz, B. P., Schmitt, A., Selvaraj, S., Pop, M., Phillippy, A. M., & Koren, S. (2019, August 21). Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLOS Computational Biology, 15(8), e1007273. https://doi.org/10.1371/journal.pcbi.1007273

Goubert, C., Craig, R. J., Bilat, A. F., Peona, V., Vogan, A. A., & Protasio, A. V. (2022, March 30). A beginner's guide to manual curation of transposable elements. Mobile DNA, 13(1). https://doi.org/10.1186/s13100-021-00259-7

Henry A, 1914. The P. nigra. Gardeners' Chronicle, 56:1-2, 46-47.

Hollender, C. A., Hill, J. L., Waite, J., & Dardick, C. (2020, April 8). Opposing influences of TAC1 and LAZY1 on Lateral Shoot Orientation in *Arabidopsis*. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-62962-4

Howard, N. P., Peace, C., Silverstein, K. A. T., Poets, A., Luby, J. J., Vanderzande, S., Durel, C. E., Muranty, H., Denancé, C., & van de Weg, E. (2021, September 1). The use of shared haplotype length information for pedigree reconstruction in asexually propagated outbreeing crops, demonstrated for apple and sweet cherry. Horticulture Research, 8(1). https://doi.org/10.1038/s41438-021-00637-5

IGATech / Projects & Funding | IGATech. (n.d.). IGATech / Projects & Funding | IGATech. https://igatechnology.com/igatech/current-projects/

Jurka, J., Kapitonov, V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and Genome Research, 110(1–4), 462–467. https://doi.org/10.1159/000084979

Koch, O., De Ávila, A. L., Heinen, H., & Albrecht, A. (2022). Retreat of Major European Tree Species Distribution under Climate Change—Minor Natives to the Rescue? Sustainability, 14(9), 5213. https://doi.org/10.3390/su14095213

Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., Wang, Z., Rasko, D. A., McCombie, W. R., Jarvis, E. D., & Phillippy, A. M. (2012, July). Hybrid error correction and

de novo assembly of single-molecule sequencing reads. Nature Biotechnology, 30(7), 693–700. https://doi.org/10.1038/nbt.2280

Kwon, J. M., & Goate, A. M. (2000). The candidate gene approach. Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism, 24(3), 164–168.

Lewandowski, I. (2017). Biomass Production from Lignocellulosic Energy Crops. Encyclopedia of Applied Plant Sciences, 159–163. https://doi.org/10.1016/b978-0-12-394807-6.00171-4

Li, C., Xing, H., Li, C., Ren, Y., Li, H., Wan, X. Q., Lian, C., Mi, J. X., & Zhang, S. (2022, October 25). Chromosome-scale genome assembly provides insights into the molecular mechanisms of tissue development of *Populus wilsonii*. Communications Biology, 5(1). https://doi.org/10.1038/s42003-022-04106-0

Magris, G., Jurman, I., Fornasiero, A., Paparelli, E., Schwope, R., Marroni, F., Di Gaspero, G., & Morgante, M. (2021, December 21). The genomes of 204 Vitis vinifera accessions reveal the origin of European wine grapes. Nature Communications, 12(1). https://doi.org/10.1038/s41467-021-27487-y

Maloy, S., & Hughes, K. (2013, March 3). Brenner's Encyclopedia of Genetics. Academic Press.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021, July 28). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution, 38(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Map of Life - Species Maps. (n.d.). Map of Life - Species Maps. https://mol.org/species/Populus_nigra

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Molecular Ecology Resources, 17(6), 1136–1147. https://doi.org/10.1111/1755-0998.12654

Neumann, P., Novák, P., Hoštáková, N., & Macas, J. (2019, January 3). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains

and provides a reference for element classification. Mobile DNA, 10(1). https://doi.org/10.1186/s13100-018-0144-1

Nicotra, A., Atkin, O., Bonser, S., Davidson, A., Finnegan, E., Mathesius, U., Poot, P., Purugganan, M., Richards, C., Valladares, F., & van Kleunen, M. (2010, December). Plant phenotypic plasticity in a changing climate. Trends in Plant Science, 15(12), 684–692. https://doi.org/10.1016/j.tplants.2010.09.008

OpenAI. (2024). ChatGPT [Large language model]. https://chat.openai.com

Otto, D., Petersen, R., Brauksiepe, B., Braun, P. , & Schmidt, E. R. (2013, December 15). The columnar mutation ("Co gene") of apple (*Malus × domestica*) is associated with an integration of a Gypsy-like retrotransposon. Molecular Breeding, 33(4), 863–880. https://doi.org/10.1007/s11032-013-0001-3

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2022, March 8). Author Correction: Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biology, 23(1). https://doi.org/10.1186/s13059-022-02645-7

Panchy, N., Lehti-Shiu, M. D., & Shiu, S. (2016). Evolution of gene duplication in plants. Plant Physiology, 171(4), 2294–2316. https://doi.org/10.1104/pp.16.00523

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007, October 11). Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System Sciences, 11(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Petersen, R., & Krost, C. (2013). Tracing a key player in the regulation of plant architecture: the columnar growth habit of apple trees (Malus × domestica). Planta, 238(1), 1–22. https://doi.org/10.1007/s00425-013-1898-9

Pevsner, J. (2003, October 21). Bioinformatics and Functional Genomics. Wiley-Liss. https://doi.org/10.1604/9780471210047

Pinosio, S., Giacomello, S., Faivre-Rampant, P., Taylor, G., Jorge, V., Le Paslier, M. C., Zaina, G., Bastien, C., Cattonaro, F., Marroni, F., & Morgante, M. (2016, August 7). Characterization of the Poplar Pan-Genome by Genome-Wide Identification of Structural Variation. Molecular Biology and Evolution, 33(10), 2706–2719. https://doi.org/10.1093/molbev/msw161

Populus nigra - Black poplar Range Map. (n.d.). plantmaps.com. https://www.plantmaps.com/nrm-populus-nigra-black-poplar-native-range-map '*Populus nigra*' from the website Trees and Shrubs Online (https://treesandshrubsonline.org/articles/populus/populus-nigra/). Accessed 2023-03-23

Populus nigra "Italica" - Plant Finder. (n.d.). https://www.missouribotanicalgarden.org/PlantFinder/PlantFinderDetails.aspx?taxonid=274293

Rayson, S., Arciga-Reyes, L., Wootton, L., De Torres Zabala, M., Truman, W., Graham, N., Grant, M., & Davies, B. (2012, February 22). A Role for Nonsense-Mediated mRNA Decay in Plants: Pathogen Responses Are Induced in Arabidopsis thaliana NMD Mutants. PLoS ONE, 7(2), e31917. https://doi.org/10.1371/journal.pone.0031917

Rodríguez, B. D., Galanti, D., Nunn, A., Peña-Ponton, C., Pérez-Bello, P., Sammarco, I., Jandrasits, K., Becker, C., De Paoli, E., Verhoeven, K. J., Opgenoorth, L., & Heer, K. (2022, November 17). Epigenetic variation in the Lombardy poplar along climatic gradients is independent of genetic structure and persists across clonal reproduction. https://doi.org/10.1101/2022.11.17.516862

Ruan, J., & Li, H. (2019, December 9). Fast and accurate long-read assembly with wtdbg2. Nature Methods, 17(2), 155–158. https://doi.org/10.1038/s41592-019-0669-3

Rushforth, K. (1999). Trees of Britain and Europe. ISBN 0-00-220013-9

Sahebi, M., Hanafi, M. M., van Wijnen, A. J., Rice, D., Rafii, M., Azizi, P., Osman, M., Taheri, S., Bakar, M. F. A., Isa, M. N. M., & Noor, Y. M. (2018, July). Contribution of transposable elements in the plant's genome. Gene, 665, 155–166. https://doi.org/10.1016/j.gene.2018.04.050

Sakamoto, T., Morinaka, Y., Ohnishi, T., Sunohara, H., Fujioka, S., Ueguchi-Tanaka, M., Mizutani, M., Sakata, K., Takatsuto, S., Yoshida, S., Tanaka, H., Kitano, H., & Matsuoka, M. (2005, December 20). Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice. Nature Biotechnology, 24(1), 105–109. https://doi.org/10.1038/nbt1173

San-Miguel-Ayanz, J., De Rigo, D., Caudullo, G., Durrant, T. H., & Mauri, A. (2016). European Atlas of Forest Tree Species.

Scaglione, D., Pinosio, S., Marroni, F., Di Centa, E., Fornasiero, A., Magris, G., Scalabrin, S., Cattonaro, F., Taylor, G., & Morgante, M. (2019, March 30). Single primer enrichment technology as a tool for massive genotyping: a benchmark on black poplar and maize. Annals of Botany, 124(4), 543–551. https://doi.org/10.1093/aob/mcz054

Smit, Hubley, & Green. (n.d.). RepeatMasker Home Page. Retrieved September 29, 2023, from http://www.repeatmasker.org/

Smulders, M., Cottrell, J., Lefèvre, F., van der Schoot, J., Arens, P., Vosman, B., Tabbener, H., Grassi, F., Fossati, T., Castiglione, S., Krystufek, V., Fluch, S., Burg, K., Vornam, B., Pohl, A., Gebhardt, K., Alba, N., Agúndez, D., Maestro, C., . . . Boerjan, W. (2008, April). Structure of the genetic diversity in black poplar (*Populus nigra L.*) populations across European river systems: Consequences for conservation and restoration. Forest Ecology and Management, 255(5–6), 1388–1399. https://doi.org/10.1016/j.foreco.2007.10.063

Storer, J., Hubley, R., Rosen, J., & Smit, A. (2022, April 17). Methodologies for the De novo Discovery of Transposable Element Families. Genes, 13(4), 709. https://doi.org/10.3390/genes13040709

Taniguchi, M., Furutani, M., Nishimura, T., Nakamura, M., Fushita, T., Iijima, K., Baba, K., Tanaka, H., Toyota, M., Tasaka, M., & Morita, M. T. (2017, August). The *Arabidopsis* LAZY1 Family Plays a Key Role in Gravity Signaling within Statocytes and in Branch Angle Control of Roots and Shoots. The Plant Cell, 29(8), 1984–1999. https://doi.org/10.1105/tpc.16.00575

The Introduction to Genetic Diversity, hosted by the National Biological Information Infrastructure and preserved by the U.S. Geological Survey, was archived on February 25, 2011, and can be referenced as a valuable source.

Thomas LK, Tölle L, Ziegenhagen B, Leyer I, 2012. Are vegetative reproduction capacities the cause of widesthe pread invasion of Eurasian Salicaceae in Patagonian river landscapes? PLoS ONE, 7(12):e50652 http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0050652

Tobutt, K. (1985, July). BREEDING COLUMNAR APPLES AT EAST MALLING. Acta Horticulturae, 159, 63–68. https://doi.org/10.17660/actahortic.1985.159.8

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., . . . Rokhsar, D. (2006, September 15). The Genome of Black Cottonwood, *Populus trichocarp*a (Torr. & Gray). Science, 313(5793), 1596–1604. https://doi.org/10.1126/science.1128691

Vanden Broeck, A., Cox, K., Van Braeckel, A., Neyrinck, S., De Regge, N., & Van Looy, K. (2021). Reintroduced Native *Populus nigra* in Restored Floodplain Reduces Spread of Exotic Poplar Species. Frontiers in Plant Science, 11. https://doi.org/10.3389/fpls.2020.580653

Wright, S. (1931, May 1). EVOLUTION IN MENDELIAN POPULATIONS. Genetics, 16(3), 290–290. https://doi.org/10.1093/genetics/16.3.290

Zhang, Z., Chen, Y., Zhang, J., Ma, X., Li, Y., Li, M., Wang, D., Kang, M., Wu, H., Yang, Y., Olson, M. S., DiFazio, S. P., Wan, D., Liu, J., & Ma, T. (2020, March 4). Improved genome assembly provides new insights into genome evolution in a desert poplar (*Populus euphratica*). Molecular Ecology Resources, 20(3), 781–794. https://doi.org/10.1111/1755-0998.13142

Zhao, Y., Li, X., Xie, J., Xu, W., Chen, S., Zhang, X., Liu, S., Wu, J., El-Kassaby, Y. A., & Zhang, D. (2022, February 1). Transposable Elements: Distribution, Polymorphism, and Climate Adaptation in *Populus*. Frontiers in Plant Science, 13. https://doi.org/10.3389/fpls.2022.814718

Zhu, M., & Zhao, S. (2007). Candidate Gene Identification Approach: Progress and Challenges. International Journal of Biological Sciences, 420–427. https://doi.org/10.7150/ijbs.3.420

Zsuffa, L. (1974). The genetics of *Populus nigra* L. Academia scientiarum et artium slavorum meridionalium (Zagreb). Ann. For. 6, 29–53.

10 APPENDIX

10.1 APPENDIX I

EpiDiverse Collection: Columnar P. nigra trees from Europe					
Sample	**Country**	**Site**	**Community Type**	**Latitude**	**Longitude**
PNI_BG_01	Marburg botanical garden			±50.804634	±8.810759
PNI_BG_02	Marburg botanical garden			±50.804635	±8.810760
PNI_BG_03	Marburg botanical garden			±50.804636	±8.810761
PNI_02_01	Spain	NA	city edge	41.60239303	-4.756418997
PNI_02_02	Spain	NA	city edge	41.60245397	-4.756910009
PNI_02_03	Spain	NA	city edge	41.60249798	-4.757227013
PNI_02_04	Spain	NA	city edge	41.60248096	-4.757904019
PNI_02_05	Spain	NA	village square	41.55542403	-4.790110979
PNI_02_06	Spain	NA	village square	41.55536503	-4.790136963
PNI_02_07	Spain	NA	village square	41.55534801	-4.790190021
PNI_02_08	Spain	NA	village square	41.55540501	-4.790236037
PNI_02_09	Spain	NA	village edge	41.52642097	-4.712072024
PNI_02_10	Spain	NA	village edge	41.52939797	-4.705647966

PNI_02_11	Spain	NA	village edge	41.52935699	-4.706132021
PNI_02_12	Spain	NA	tech park	41.522959	-4.714975012
PNI_02_15	Spain	NA	rural area	41.76442201	-4.698034013
PNI_02_16	Spain	NA	rural area	41.70664304	-4.784212969
PNI_02_17	Spain	NA	village edge	41.677413	-4.917102009
PNI_02_18	Spain	NA	rural area	41.67213399	-4.914265992
PNI_02_19	Spain	NA	rural area	41.67206501	-4.914234979
PNI_02_20	Spain	NA	rural area	41.67202301	-4.914428014
PNI_02_21	Spain	NA	rural area	41.67201103	-4.914478976
PNI_02_22	Spain	NA	village edge	41.58614002	-4.811592037
PNI_02_23	Spain	NA	village edge	41.58605101	-4.811792029
PNI_02_24	Spain	NA	village edge	41.58525799	-4.812767012
PNI_02_25	Spain	NA	village edge	41.58523	-4.812819986
PNI_02_26	Spain	NA	village edge	41.585201	-4.812912019
PNI_03_01	France	NA	village	44.70526303	-0.365649965

PNI_03_02	France	NA	village	44.70530502	-0.365614006
PNI_03_03	France	NA	village	44.70526999	-0.365662035
PNI_03_04	France	NA	rural area	44.569556	0.320233041
PNI_03_05	France	NA	rural area	44.56955197	0.320331026
PNI_03_06	France	NA	rural area	44.56998297	0.324509991
PNI_03_07	France	NA	rural area	44.57002203	0.324418042
PNI_03_08	France	NA	rural area	44.57395096	0.329317013
PNI_03_09	France	NA	village	44.59701	0.358749982
PNI_03_10	France	NA	village	44.59694596	0.358591983
PNI_03_11	France	NA	village	44.59691503	0.358547978
PNI_03_12	France	NA	village	44.59693297	0.358464997
PNI_03_13	France	NA	rural area	44.63037601	0.356697002
PNI_03_14	France	NA	rural area	44.63028297	0.357145015
PNI_03_15	France	NA	rural area	44.63048397	0.356859025
PNI_03_16	France	NA	rural area	44.63045103	0.356862964
PNI_03_17	France	NA	rural area	44.67537903	0.320865959
PNI_03_18	France	NA	rural area	44.720271	0.278454032
PNI_03_19	France	NA	rural area	44.72023102	0.278499965
PNI_03_20	France	NA	rural area	44.72023303	0.278515974
PNI_03_21	France	NA	rural area	44.72019698	0.278518992
PNI_03_22	France	NA	rural area	44.627771	0.193083007
PNI_03_23	France	NA	rural area	44.6278	0.193150984

PNI_03_24	France	NA	rural area	44.62772598	0.193156013
PNI_03_25	France	NA	rural area	44.62700204	0.193091976
PNI_03_26	France	NA	rural area	44.64336896	0.171537995
PNI_03_27	France	NA	rural area	44.64345203	0.171677973
PNI_03_28	France	NA	rural area	44.64340199	0.171810994
PNI_03_29	France	NA	village	44.64623298	0.148626985
PNI_03_30	France	NA	village	44.64626097	0.148661016
PNI_03_31	France	NA	rural area	44.65780796	0.115119992
PNI_03_32	France	NA	rural area	44.65767997	0.115146982
PNI_03_33	France	NA	rural area	44.657901	0.11511798
PNI_03_34	France	NA	rural area	44.65788097	0.115112029
PNI_03_35	France	NA	rural area	44.61576803	0.097420011
PNI_03_36	France	NA	rural area	44.61584698	0.09731004
PNI_03_37	France	NA	rural area	44.61584296	0.097280033
PNI_03_38	France	NA	rural area	44.61596802	0.097192023
PNI_03_39	France	NA	rural area	44.82337402	-0.023662029
PNI_03_40	France	NA	rural area	44.82336598	-0.023596985
PNI_03_41	France	NA	rural area	44.823371	-0.023550969
PNI_03_42	France	NA	rural area	44.82334703	-0.023526996
PNI_03_43	France	NA	rural area	44.82334603	-0.023462037

PNI_03_44	France	NA	rural area	44.82471999	-0.003649984
PNI_03_45	France	NA	rural area	44.82467498	-0.003674962
PNI_03_46	France	NA	rural area	44.82464698	-0.003697006
PNI_03_47	France	NA	rural area	44.82462502	-0.003683008
PNI_03_48	France	NA	rural area	44.53446298	0.351552023
PNI_03_49	France	NA	rural area	44.53448301	0.351497037
PNI_03_50	France	NA	rural area	44.53449097	0.351382038
PNI_03_51	France	NA	rural area	44.53447999	0.351347001
PNI_03_52	France	NA	rural area	44.53448603	0.351325963
PNI_03_53	France	NA	rural area	44.51847198	0.413812967
PNI_03_54	France	NA	rural area	44.51849503	0.413757982
PNI_03_55	France	NA	rural area	44.518517	0.413751025
PNI_03_56	France	NA	rural area	44.518446	0.413692016
PNI_03_57	France	NA	rural area	44.82337	-0.02366
PNI_03_58	France	NA	rural area	44.82337	-0.0236
PNI_03_59	France	NA	rural area	44.82337	-0.02355
PNI_03_60	France	NA	rural area	44.82335	-0.02353
PNI_04_01	France	NA	garden	43.49526004	6.508860961
PNI_04_02	France	NA	garden	43.49524504	6.50890396
PNI_04_03	France	NA	garden	43.49518703	6.508956011
PNI_04_04	France	NA	garden	43.49521897	6.508981995

PNI_04_05	France	NA	garden	44.03819002	6.055751033
PNI_04_06	France	NA	garden	44.03881498	6.054928014
PNI_04_07	France	NA	garden	44.03898597	6.05499398
PNI_04_08	France	NA	garden	44.03898002	6.055013007
PNI_04_09	France	NA	highway exit	44.03960204	5.964314025
PNI_04_10	France	NA	highway exit	44.03960204	5.964333974
PNI_04_11	France	NA	village	43.99195603	5.046654968
PNI_04_12	France	NA	village	43.991994	5.04662496
PNI_04_13	France	NA	village	43.99192703	5.046578022
PNI_04_14	France	NA	village	43.99192502	5.046521025
PNI_04_15	France	NA	rural area	43.989434	5.023323018
PNI_04_16	France	NA	rural area	43.98945102	5.023252023
PNI_04_17	France	NA	rural area	43.98946702	5.023265015
PNI_04_18	France	NA	rural area	43.98946401	5.023192009
PNI_04_19	France	NA	rural area	43.989463	5.023157978
PNI_04_20	France	NA	rural area	43.97419604	5.00434597
PNI_04_21	France	NA	rural area	43.974146	5.004387042
PNI_04_22	France	NA	rural area	43.97409897	5.004410008
PNI_04_23	France	NA	rural area	43.97405396	5.004419982
PNI_04_24	France	NA	rural area	43.97401297	5.004455019
PNI_04_25	France	NA	rural area	43.87937098	5.005140994
PNI_04_26	France	NA	rural area	43.87935396	5.005087014
PNI_04_27	France	NA	rural area	43.879375	5.005106963

PNI_04_28	France	NA	rural area	43.87933603	5.005146023
PNI_04_29	France	NA	rural area	43.87935497	5.005067987
PNI_04_30	France	NA	rural area	43.91936498	5.007694038
PNI_04_31	France	NA	rural area	43.91929097	5.007148962
PNI_04_32	France	NA	rural area	43.91922701	5.006736992
PNI_04_33	France	NA	rural area	43.96380097	4.995710012
PNI_04_34	France	NA	rural area	43.96382201	4.995790981
PNI_04_35	France	NA	rural area	43.96381496	4.995838003
PNI_04_36	France	NA	rural area	43.96376702	4.995974964
PNI_04_37	France	NA	rural area	43.98372098	4.956642967
PNI_04_38	France	NA	rural area	43.98375199	4.956713961
PNI_04_39	France	NA	rural area	43.98377303	4.956737012
PNI_04_40	France	NA	rural area	43.98378904	4.956773976
PNI_04_41	France	NA	rural area	43.98377596	4.956804989
PNI_04_42	France	NA	rural area	43.84862297	4.914631024
PNI_04_43	France	NA	rural area	43.84861501	4.914573021
PNI_04_44	France	NA	rural area	43.84859296	4.914493986
PNI_04_45	France	NA	rural area	43.84859598	4.914456016
PNI_04_46	France	NA	rural area	43.84857704	4.914352996
PNI_04_47	France	NA	rural area	43.83894799	4.875743017
PNI_04_48	France	NA	rural area	43.83894204	4.875762966
PNI_04_49	France	NA	rural area	43.83892402	4.875948038
PNI_04_50	France	NA	rural area	43.83891698	4.876029007

PNI_04_51	France	NA	rural area	44.01302897	5.014039977
PNI_04_52	France	NA	rural area	44.01305102	5.014003012
PNI_04_53	France	NA	rural area	44.01311103	5.014020028
PNI_04_54	France	NA	rural area	43.95167897	4.989536991
PNI_04_55	France	NA	rural area	43.95157998	4.989539003
PNI_04_56	France	NA	rural area	43.95154696	4.989407994
PNI_05_01	France	NA	rural area	46.12555498	3.396397959
PNI_05_02	France	NA	rural area	46.12554098	3.396323025
PNI_05_03	France	NA	rural area	46.12544702	3.396184975
PNI_05_04	France	NA	rural area	46.12511703	3.394886032
PNI_05_05	France	NA	rural area	46.12506598	3.394767009
PNI_05_06	France	NA	rural area	46.18426301	3.487402964
PNI_05_07	France	NA	rural area	46.18425102	3.487378992
PNI_05_08	France	NA	rural area	46.18426703	3.487353008
PNI_05_09	France	NA	rural area	46.18440399	3.487171037
PNI_05_10	France	NA	rural area	46.18472602	3.497196967
PNI_05_11	France	NA	rural area	46.18470901	3.497181041
PNI_05_12	France	NA	rural area	46.18468998	3.497191016
PNI_05_13	France	NA	rural area	46.18467598	3.497195961
PNI_05_14	France	NA	rural area	46.18456601	3.497193027
PNI_05_15	France	NA	rural area	46.12678796	3.342815973
PNI_05_16	France	NA	rural area	46.12671202	3.342751013
PNI_05_17	France	NA	village	46.13019797	3.324070014

PNI_05_18	France	NA	village	46.1301579	3.32409801
PNI_05_19	France	NA	village	46.1301220 3	3.32411502 5
PNI_05_20	France	NA	village	46.130175	3.32409801
PNI_05_21	France	NA	village	46.1301719 9	3.32416397 5
PNI_05_22	France	NA	rural area	46.1294460 3	3.27876498 9
PNI_05_23	France	NA	rural area	46.1294000 1	3.27872500 8
PNI_05_24	France	NA	rural area	46.1294359 7	3.27868301 4
PNI_05_25	France	NA	rural area	46.1293440 2	3.27861595 9
PNI_05_26	France	NA	rural area	46.1293519 8	3.27858301 8
PNI_05_27	France	NA	rural area/village	46.1401079 8	3.27030496 7
PNI_05_28	France	NA	rural area/village	46.1400850 1	3.27027898 3
PNI_05_29	France	NA	rural area/village	46.1399669 9	3.27022902 7
PNI_05_30	France	NA	rural area/village	46.1398349 8	3.27001696 5
PNI_05_31	France	NA	rural area/village	46.1398349 8	3.26997103 2
PNI_05_32	France	NA	rural area	46.1612759 7	3.28917103 8
PNI_05_33	France	NA	rural area	46.1613310 4	3.28915704
PNI_05_34	France	NA	rural area	46.1619959 8	3.28935803 8
PNI_05_35	France	NA	rural area	46.1620379 7	3.28932400 8
PNI_05_36	France	NA	rural area	46.1620659 7	3.28928201 5
PNI_05_37	France	NA	rural area	46.1900669 7	3.34207702 4
PNI_05_38	France	NA	rural area	46.1900419 9	3.34214600 7
PNI_05_39	France	NA	horse paddock	46.1900089 7	3.34219596 3
PNI_05_40	France	NA	horse paddock	46.1899870 1	3.34226402 5

PNI_05_41	France	NA	horse paddock	46.18995499	3.342284979
PNI_05_42	France	NA	village	46.19085202	3.405578993
PNI_05_43	France	NA	parking lot	46.14118103	3.396171983
PNI_05_44	France	NA	parking lot	46.14118002	3.396149017
PNI_05_45	France	NA	parking lot	46.14108698	3.396126973
PNI_05_46	France	NA	parking lot	46.14105798	3.396153962
PNI_05_47	France	NA	parking lot	46.14105103	3.396202996
PNI_05_48	France	NA	village	46.15707597	3.460442992
PNI_05_49	France	NA	village	46.20435904	3.472590968
PNI_05_50	France	NA	village	46.20434898	3.472606977
PNI_05_51	France	NA	village	46.20432501	3.472631034
PNI_05_52	France	NA	village	46.20428603	3.472622987
PNI_05_53	France	NA	village	46.20426902	3.472636985
PNI_05_54	France	NA	rural area	46.24041799	3.527185991
PNI_05_55	France	NA	rural area	46.240335	3.527338961
PNI_05_56	France	NA	rural area	46.242218	3.527987972
PNI_06_01	Germany	NA	rural area	49.39818299	8.368963022
PNI_06_02	Germany	NA	village	49.398098	8.368882975
PNI_06_03	Germany	NA	village	49.39765602	8.368521966
PNI_06_04	Germany	NA	village	49.39751504	8.368437979
PNI_06_05	Germany	NA	village	49.39684499	8.367826017
PNI_06_06	Germany	NA	village/rural area	49.37368898	8.303722981
PNI_06_07	Germany	NA	village/rural area	49.36354604	8.299838975

PNI_06_08	Germany	NA	rural area	49.36323004	8.29344199
PNI_06_09	Germany	NA	rural area	49.36320196	8.293333026
PNI_06_10	Germany	NA	rural area	49.36318997	8.293261025
PNI_06_11	Germany	NA	village/rural area	49.37382703	8.304167977
PNI_06_12	Germany	NA	village/rural area	49.37376098	8.304066975
PNI_06_13	Germany	NA	village/rural area	49.37353701	8.303842004
PNI_06_14	Germany	NA	village/rural area	49.37346401	8.303788025
PNI_06_15	Germany	NA	village	49.54264602	8.226536969
PNI_06_16	Germany	NA	soccer place	49.55122297	8.234112038
PNI_06_17	Germany	NA	soccer place	49.55117997	8.234091997
PNI_06_18	Germany	NA	soccer place	49.55113697	8.234109012
PNI_06_19	Germany	NA	soccer place	49.55113596	8.234111024
PNI_06_20	Germany	NA	soccer place	49.55089599	8.234174978
PNI_06_21	Germany	NA	rural area	49.515808	8.297568988
PNI_06_22	Germany	NA	rural area	49.515937	8.297882974
PNI_06_23	Germany	NA	rural area	49.51585603	8.297783984
PNI_06_24	Germany	NA	rural area	49.51589501	8.298005015
PNI_06_25	Germany	NA	rural area	49.51595603	8.298240965
PNI_06_26	Germany	NA	village	49.50705897	8.301502028
PNI_06_27	Germany	NA	village	49.50706199	8.301287033
PNI_06_28	Germany	NA	village	49.50701203	8.300989978
PNI_06_29	Germany	NA	village	49.50699602	8.30065202
PNI_06_30	Germany	NA	village edge	49.465172	8.294812012

PNI_06_31	Germany	NA	village edge	49.46531801	8.295113007
PNI_06_32	Germany	NA	village edge	49.46548398	8.294659965
PNI_06_33	Germany	NA	village edge	49.46554198	8.294745963
PNI_06_34	Germany	NA	village edge	49.46562303	8.294587964
PNI_06_35	Germany	NA	village park	49.70327698	8.250559
PNI_06_36	Germany	NA	village park	49.70331101	8.250758993
PNI_06_37	Germany	NA	village park	49.70329198	8.250730997
PNI_06_38	Germany	NA	village park	49.70342701	8.250752036
PNI_06_39	Germany	NA	village park	49.70337798	8.250906011
PNI_06_40	Germany	NA	village	49.67607896	8.290893976
PNI_06_41	Germany	NA	village	49.60849701	8.180084964
PNI_06_42	Germany	NA	village	49.60835997	8.17999796
PNI_06_43	Germany	NA	village	49.60836198	8.17977299
PNI_06_44	Germany	NA	village	49.60832099	8.179683974
PNI_06_45	Germany	NA	rural area	49.58559698	8.193011032
PNI_06_46	Germany	NA	rural area	49.58569204	8.193261987
PNI_06_47	Germany	NA	rural area	49.58569204	8.193356032
PNI_06_48	Germany	NA	rural area	49.585714	8.193515036
PNI_06_49	Germany	NA	rural area	49.491979	8.602501033
PNI_06_50	Germany	NA	rural area	49.49194799	8.60275601
PNI_06_51	Germany	NA	rural area	49.49192796	8.603143003
PNI_06_52	Germany	NA	rural area	49.49201404	8.603018029
PNI_06_53	Germany	NA	rural area	49.51731398	8.644676004

PNI_06_54	Germany	NA	rural area	49.517289	8.644653959
PNI_06_55	Germany	NA	rural area	49.59179599	8.615541011
PNI_06_56	Germany	NA	rural area	49.58763496	8.618565034
PNI_07_01	Germany	Leegebruch	meadow	52.72628	13.11432
PNI_07_02	Germany	Marwitz Ost	meadow	52.70788	13.13274
PNI_07_03	Germany	Marwitz Ost	meadow	52.70795	13.13266
PNI_07_04	Germany	Marwitz Ost	meadow	52.70799	13.13265
PNI_07_05	Germany	Marwitz Ost	meadow	52.70805	13.1326
PNI_07_06	Germany	Marwitz Ost	meadow	52.70817	13.13256
PNI_07_07	Germany	Altruppin	village	52.72904	13.07268
PNI_07_08	Germany	Altruppin	village	52.72913	13.07267
PNI_07_09	Germany	Altruppin	village	52.72928	13.07263
PNI_07_10	Germany	Altruppin	village	52.73007	13.0725
PNI_07_11	Germany	Altruppin	meadow	52.921	12.86119
PNI_07_12	Germany	Altruppin	meadow	52.92096	12.86122
PNI_07_13	Germany	Altruppin	meadow	52.92089	12.86126
PNI_07_14	Germany	Altruppin	meadow	52.92085	12.86128
PNI_07_15	Germany	Altruppin	meadow	52.9208	12.86133
PNI_07_16	Germany	Altruppin	meadow	52.90407	12.87721
PNI_07_17	Germany	Altruppin	meadow	52.90413	12.87715
PNI_07_18	Germany	Altruppin	meadow	52.90427	12.87707
PNI_07_19	Germany	Altruppin	meadow	52.90428	12.87701
PNI_07_20	Germany	Altruppin	meadow	52.90379	12.87673

PNI_07_21	Germany	Neuruppin	village	52.89022	12.80879
PNI_07_22	Germany	Neuruppin	village	52.89024	12.80872
PNI_07_23	Germany	Neuruppin	village	52.89023	12.80871
PNI_07_24	Germany	Neuruppin	village	52.89018	12.80887
PNI_07_25	Germany	Neuruppin	village	52.89013	12.80892
PNI_07_26	Germany	Zehdenik	village	52.98043	13.35014
PNI_07_27	Germany	Zehdenik	village	52.98048	13.35
PNI_07_28	Germany	Zehdenik	village	52.98049	13.35001
PNI_07_29	Germany	Zehdenik	village	52.98065	13.35002
PNI_07_30	Germany	Zehdenik West	meadow	52.98497	13.29543
PNI_07_31	Germany	Zehdenik West	meadow	52.98496	13.29543
PNI_07_32	Germany	Zehdenik West	meadow	52.98483	13.29539
PNI_07_33	Germany	Zehdenik West	meadow	52.98476	13.29537
PNI_07_34	Germany	Zehdenik West	meadow	52.9847	13.29533
PNI_07_35	Germany	Neuruppin	village	52.92964	12.79879
PNI_07_36	Germany	Neuruppin	village	52.92961	12.79873
PNI_07_37	Germany	Neuruppin	village	52.92959	12.79877
PNI_07_38	Germany	Neuruppin	village	52.92947	12.79872
PNI_07_39	Germany	Paulinenaue	meadow	52.65886	12.80706
PNI_07_40	Germany	Paulinenaue	meadow	52.65874	12.80718
PNI_07_41	Germany	Nauen	meadow	52.61397	12.87025
PNI_07_42	Germany	Nauen	meadow	52.61397	12.87005
PNI_07_43	Germany	Nauen	village	52.61227	12.87764

PNI_07_44	Germany	Kiemberg	village	52.66299	12.9126
PNI_07_45	Germany	Kiemberg	village	52.66302	12.9126
PNI_07_46	Germany	Kiemberg	village	52.66301	12.91267
PNI_07_47	Germany	Kiemberg	village	52.66304	12.91265
PNI_07_48	Germany	Kiemberg	village	52.66309	12.91265
PNI_07_49	Germany	Elstal	village	52.54684	12.98732
PNI_07_50	Germany	Elstal	village	52.547	12.98682
PNI_07_51	Germany	Elstal	village	52.54672	12.98893
PNI_07_52	Germany	Elstal	village	52.54672	12.98904
PNI_07_53	Germany	Elstal	village	52.54667	12.98936
PNI-09-01	Norway	Oslo Botanical Garden	Botanical Garden	59.919535	10.772283
PNI-09-02	Norway	Oslo Botanical Garden	Next to House	59.919535	10.772283
PNI-09-03	Norway	Oslo Botanical Garden	Next to House	59.919535	10.772283
PNI-09-04	Norway	Oslo Botanical Garden	Next to House	59.919535	10.772283
PNI-09-07	Norway	Moss	Next to House/Road	59.433983	10.655833
PNI-09-08	Norway	Moss	Next to House/Road	59.433983	10.655833
PNI-09-09	Norway	Moss	Next to House/Road	59.433983	10.655833
PNI-09-10	Norway	Moss	Next to House/Road	59.433983	10.655833
PNI-09-11	Norway	Moss	Next to House/Road	59.433983	10.655833
PNI-09-12	Norway	Moss	Next to House/Road	59.433983	10.655833
PNI-09-13	Norway	Moss	Next to House/Road	59.433983	10.655833
PNI-09-14	Norway	Vassum Tunnelen	turnaround	59.70685	10.655833
PNI-09-17	Norway	Oslo, Hoffsveien		59.935417	10.680583
PNI-09-18	Norway	Oslo, Hoffsveien		59.935417	10.680650
PNI-09-19	Norway	Oslo, Hoffsveien		59.935467	10.680833

PNI-09-20	Norway	Oslo, Hoffsveien		NA	NA
PNI-09-21	Norway	Oslo, Hoffsveien		59.935483	10.680900
PNI-09-22	Norway	Oslo, Hoffsveien		59.935500	10.680833
PNI-09-23	Norway	Oslo, Hoffsveien		59.935467	10.680683
PNI-09-24	Norway	Oslo, Hoffsveien		59.935600	10.680883
PNI-09-25	Norway	Oslo, Hoffsveien		59.935583	10.680717
PNI-09-26	Norway	Oslo, Hoffsveien		59.934600	10.678350
PNI-09-27	Norway	Oslo, Hoffsveien		59.209729	10.946492
PNI-09-32	Norway	Fredrikstad	Parking lot	59.209702	10.947883
PNI-09-33	Norway	Fredrikstad	Parking lot	59.209702	10.947883
PNI-09-34	Norway	Fredrikstad	Parking lot	59.209795	10.947717
PNI-09-35	Norway	Fredrikstad	Parking lot	59.209891	10.9475
PNI-09-36	Norway	Fredrikstad	Parking lot	59.209911	10.947472
PNI-09-37	Norway	Fredrikstad	Parking lot	59.209952	10.947387
PNI-09-38	Norway	Fredrikstad	Parking lot	59.210007	10.947303
PNI-09-39	Norway	Fredrikstad	Parking lot	59.210075	10.947176
PNI-09-40	Norway	Fredrikstad	Parking lot	59.210153	10.947135
PNI-09-41	Norway	Fredrikstad	Parking lot	59.210147	10.947174
PNI-09-42	Norway	Fredrikstad	Parking lot	59.210147	10.947174
PNI-09-44	Norway	Fredrikstad	Parking lot	59.209537	10.94757
PNI-09-45	Norway	Fredrikstad	Parking lot	59.209738	10.946826
PNI-09-46	Norway	Fredrikstad	Parking lot	59.209713	10.946785
PNI-09-47	Norway	Fredrikstad	Parking lot	59.209746	10.946742
PNI-09-48	Norway	Fredrikstad	Parking lot	59.209737	10.946589
PNI-09-49	Norway	Fredrikstad	Parking lot	59.209729	10.946495
PNI-09-50	Norway	Fredrikstad	Petrol station	59.210254	10.946183
PNI-09-52	Norway	Mossveien Kjölberg Herregard	Allee	59.257100	10.894067
PNI-09-53	Norway	Mossveien Kjölberg Herregard	Allee	59.257100	10.894067
PNI-09-54	Norway	Mossveien Kjölberg Herregard	Allee	59.257083	10.894083
PNI-09-55	Norway	Mossveien Kjölberg Herregard	Allee	59.257133	10.894117
PNI-09-56	Norway	Mossveien Kjölberg Herregard	Allee	59.257117	10.894200
PNI-09-57	Norway	Mossveien Kjölberg Herregard	Allee	59.257167	10.894367
PNI-09-58	Norway	Mossveien Kjölberg Herregard	Allee	59.257167	10.894450

PNI-09-59	Norway	Mossveien Kjölberg Herregard	Allee	59.257167	10.894583
PNI-09-60	Norway	Mossveien Kjölberg Herregard	Allee	59.257183	10.894650
PNI-09-61	Norway	Mossveien Kjölberg Herregard	Allee	59.257183	10.894733
PNI-09-62	Norway	Mossveien Kjölberg Herregard	Allee	59.257117	10.894733
PNI-09-63	Norway	Mossveien Kjölberg Herregard	Allee	59.257117	10.894817
PNI-09-64	Norway	Mossveien Kjölberg Herregard	Allee	59.257133	10.895450
PNI-09-65	Norway	Mossveien Kjölberg Herregard	Allee	59.257133	10.895667
PNI-09-66	Norway	Mossveien Kjölberg Herregard	Allee	59.257183	10.895733
PNI-09-67	Norway	Mossveien Kjölberg Herregard	Allee	59.257133	10.895950
PNI-09-68	Norway	Mossveien Kjölberg Herregard	Allee	59.257183	10.896017
PNI_10_01	Italy	Ortucchio	rural area	41.96431	13.64915
PNI_10_02	Italy	Ortucchio	rural area	41.96431	13.64915
PNI_10_03	Italy	Ortucchio	rural area	41.96431	13.64915
PNI_10_04	Italy	Ortucchio	rural area	41.96431	13.64915
PNI_10_05	Italy	Ortucchio	rural area	41.96431	13.64915
PNI_10_06	Italy	Burg vio Nuevo	rural area	42.01758	13.44744
PNI_10_07	Italy	Burg vio Nuevo	rural area	42.01056	13.45409
PNI_10_08	Italy	Burg vio Nuevo	rural area	42.00847	13.45412

PNI_10_09	Italy	Borgo Ottomila	rural area	41.99184	13.51653
PNI_10_10	Italy	Borgo Ottomila	rural area	41.99184	13.51653
PNI_10_11	Italy	Ortucchio	rural area	41.95965	13.65885
PNI_10_12	Italy	Ortucchio	soccer place	41.95745	13.64759
PNI_10_13	Italy	Ortucchio	soccer place	41.95745	13.64759
PNI_10_14	Italy	Ortucchio	soccer place	41.95745	13.64759
PNI_10_15	Italy	Ortucchio	soccer place	41.95745	13.64759
PNI_10_16	Italy	Ortucchio	village	41.95591	13.64156
PNI_10_17	Italy	Ortucchio	village	41.95591	13.64156
PNI_10_18	Italy	Ortucchio	village	41.95591	13.64156
PNI_10_19	Italy	Ortucchio	village	41.95591	13.64156
PNI_10_20	Italy	Ortucchio	village	41.95591	13.64156
PNI_10_21	Italy	Trasacco	rural area	41.96325	13.52354
PNI_10_22	Italy	Luco dei Marsi	little village	41.9534	13.48916
PNI_10_23	Italy	Luco dei Marsi	little village	41.9534	13.48916
PNI_10_24	Italy	Luco dei Marsi	little village	41.9534	13.48916
PNI_10_25	Italy	Luco dei Marsi	little village	41.9534	13.48916
PNI_10_26	Italy	Luco dei Marsi	little village	41.9534	13.48916
PNI_10_27	Italy	Luco dei Marsi	little village	41.95197	13.49376
PNI_10_28	Italy	Case Incile	rural area	42.00066	13.44600
PNI_10_29	Italy	Case Incile	rural area	42.00326	13.44500
PNI_10_30	Italy	Burgo via Nuova	rural area	42.02262	13.44936
PNI_10_31	Italy	Burgo via Nuova	rural area	42.02262	13.44936

PNI_10_32	Italy	Paterno	village	42.05369	13.49077
PNI_10_33	Italy	Paterno	village	42.05369	13.49077
PNI_10_34	Italy	Cerchio Süden	rural area	42.02983	13.59477
PNI_10_35	Italy	Cerchio Süden	rural area	42.03091	13.59954
PNI_10_36	Italy	Pescina	village	42.02804	13.64304
PNI_10_37	Italy	San Pellino Süden	rural area	42.04721	13.47608
PNI_10_38	Italy	Case Incile	village	42.00148	13.44135
PNI_10_39	Italy	Case Incile	village	42.00153	13.44135
PNI_10_40	Italy	Case Incile	village	42.00156	13.44134
PNI_10_41	Italy	Case Incile	village	42.00167	13.44131
PNI_10_42	Italy	Luco dei Marsi	little village	41.94969	13.49890
PNI_10_43	Italy	Luco dei Marsi	little village	41.94973	13.49885
PNI_10_44	Italy	Luco dei Marsi	village	41.95673	13.48231
PNI_10_45	Italy	Case Incile	village	42.00126	13.43743
PNI_10_46	Italy	Case Incile	village	42.00238	13.44038
PNI_10_47	Italy	Case Incile	village	42.00237	13.44036
PNI_10_48	Italy	Luco dei Marsi	village	41.95018	13.49841
PNI_10_49	Italy	Luco dei Marsi	village	41.95024	13.49838
PNI_10_50	Italy	Luco dei Marsi	village	41.95014	13.49851
PNI_10_51	Italy	Centro Spaziale del Fucino Süden	rural area	41.96378	13.59233
PNI_10_52	Italy	Centro Spaziale del Fucino Süden	rural area	41.96377	13.59225
PNI_10_53	Italy	San Benedeto dei Marsi	rural area	42.02955	13.60013

PNI_11_01	Italy	San Damaso	meadow	44.60752	10.97677
PNI_11_02	Italy	San Damaso	meadow	44.60751	10.97689
PNI_11_03	Italy	San Damaso	meadow	44.60746	10.97701
PNI_11_04	Italy	San Damaso	meadow	44.60743	10.97716
PNI_11_05	Italy	San Damaso	meadow	44.60739	10.97724
PNI_11_06	Italy	San Domino	village	44.5872	10.97016
PNI_11_07	Italy	San Domino	village	44.58738	10.97008
PNI_11_08	Italy	San Domino	village	44.58719	10.97018
PNI_11_09	Italy	Bazzano	village	44.50396	11.09832
PNI_11_10	Italy	Bazzano	village	44.50365	11.09820
PNI_11_11	Italy	Bazzano	village	44.50296	11.09754
PNI_11_12	Italy	Anzola dell Emilia	village	44.54099	11.20600
PNI_11_13	Italy	Anzola dell Emilia	village	44.54105	11.20603
PNI_11_14	Italy	Anzola dell Emilia	village	44.54105	11.20605
PNI_11_15	Italy	Anzola dell Emilia	village	44.54093	11.20593
PNI_11_16	Italy	Budrie	rural area	44.56592	11.19268
PNI_11_17	Italy	Budrie	rural area	44.56593	11.19281
PNI_11_18	Italy	Budrie	rural area	44.5659	11.19290
PNI_11_19	Italy	Manzolino	soccer place	44.59723	11.10194
PNI_11_20	Italy	Manzolino	soccer place	44.59724	11.10203
PNI_11_21	Italy	Manzolino	soccer place	44.59726	11.10203
PNI_11_22	Italy	Panzano	little village	44.62054	11.04160
PNI_11_23	Italy	Panzano	little village	44.62047	11.04157

PNI_11_24	Italy	Panzano	little village	44.62044	11.04155
PNI_11_25	Italy	Castelfranco Emilia	village	44.58945	11.05717
PNI_11_26	Italy	Castelfranco Emilia	village	44.58957	11.05734
PNI_11_27	Italy	Castelfranco Emilia	village	44.58967	11.05731
PNI_11_28	Italy	Castellana Tarabini	meadow	44.57183	10.87097
PNI_11_29	Italy	Castellana Tarabini	meadow	44.57172	10.87109
PNI_11_30	Italy	Castellana Tarabini	meadow	44.57205	10.87051
PNI_11_31	Italy	Santa Maria in Strada	meadow	44.5726	11.15967
PNI_11_32	Italy	Santa Maria in Strada	meadow	44.57263	11.15979
PNI_11_33	Italy	Santa Maria in Strada	meadow	44.57265	11.15988
PNI_11_34	Italy	Santa Maria in Strada	meadow	44.57273	11.15981
PNI_11_35	Italy	Santa Maria in Strada	meadow	44.57262	11.15925
PNI_11_36	Italy	Budrie	soccer place	44.59235	11.17945
PNI_11_37	Italy	Budrie	soccer place	44.59236	11.17946
PNI_11_38	Italy	Budrie	soccer place	44.59237	11.17949
PNI_11_39	Italy	Recovato Ost	little village	44.63521	11.07081
PNI_11_40	Italy	Recovato Ost	little village	44.63519	11.07075
PNI_11_41	Italy	Recovato Ost	little village	44.63513	11.07070
PNI_11_42	Italy	Modena Süd-Ost	village	44.62028	10.94291
PNI_11_43	Italy	Modena Süd-Ost	village	44.62033	10.94301
PNI_11_44	Italy	Modena Süd-Ost	village	44.62034	10.94301
PNI_11_45	Italy	Modena Süd-Ost	village	44.62036	10.94303
PNI_11_46	Italy	Castelfranco Emilia	village	44.59209	11.04889

PNI_11_47	Italy	Castelfranco Emilia	village	44.5921	11.04890
PNI_11_48	Italy	Castelfranco Emilia	village	44.59207	11.04894
PNI_11_49	Italy	Castelfranco Emilia	village	44.59203	11.04898
PNI_11_50	Italy	Nonantola	village	44.67454	11.04094
PNI_11_51	Italy	Nonantola	village	44.6745	11.04102
PNI_11_52	Italy	Nonantola	village	44.67448	11.04109
PNI_11_53	Italy	Nonantola	village	44.67443	11.04124
PNI_12_01	C. Republic	Hostivice	village	50.08496	14.22576
PNI_12_02	C. Republic	Hostivice	village	50.08497	14.22564
PNI_12_03	C. Republic	Hostivice	village	50.08513	14.22476
PNI_12_04	C. Republic	Jenec	village	50.09373	14.21458
PNI_12_05	C. Republic	Jenec	village	50.09378	14.21459
PNI_12_06	C. Republic	Jenec	village	50.09381	14.21459
PNI_12_07	C. Republic	Uhost	village	50.08419	14.13913
PNI_12_08	C. Republic	Uhost	village	50.08384	14.13930
PNI_12_09	C. Republic	Braskov Süd	meadow	50.09481	14.09275
PNI_12_10	C. Republic	Braskov Süd	meadow	50.09486	14.09274
PNI_12_11	C. Republic	Braskov Süd	meadow	50.09486	14.09263
PNI_12_12	C. Republic	Braskov Süd	meadow	50.10102	14.08333
PNI_12_13	C. Republic	Kladno	village	50.14011	14.09846
PNI_12_14	C. Republic	Kladno	village	50.12792	14.12807
PNI_12_15	C. Republic	Kladno	village	50.12797	14.12814
PNI_12_16	C. Republic	Kladno	village	50.12797	14.12820

PNI_12_17	C. Republic	Kladno	village	50.12799	14.12822
PNI_12_18	C. Republic	Kladno	village	50.128	14.12832
PNI_12_19	C. Republic	Bustehrad	rural area	50.15293	14.17706
PNI_12_20	C. Republic	Bustehrad	rural area	50.1529	14.17704
PNI_12_21	C. Republic	Bustehrad	rural area	50.15285	14.17701
PNI_12_22	C. Republic	Bustehrad	rural area	50.15253	14.17695
PNI_12_23	C. Republic	Velke Prilepy Süd	rural area	50.1563	14.30478
PNI_12_24	C. Republic	Velke Prilepy Süd	rural area	50.15127	14.32182
PNI_12_25	C. Republic	Velke Prilepy Süd	rural area	50.15146	14.32129
PNI_12_26	C. Republic	Horomerice Nord	rural area	50.14022	14.34368
PNI_12_27	C. Republic	Horomerice Nord	rural area	50.14019	14.34359
PNI_12_28	C. Republic	Horomerice Nord	rural area	50.14012	14.34350
PNI_12_29	C. Republic	Horomerice Nord	rural area	50.14006	14.34341
PNI_12_30	C. Republic	Zlicin	village	50.06614	14.28676
PNI_12_31	C. Republic	Zlicin	village	50.06615	14.28670
PNI_12_32	C. Republic	Zlicin	village	50.06691	14.28611
PNI_12_33	C. Republic	Zlicin	village	50.06689	14.28606
PNI_12_34	C. Republic	Rudna	village	50.03555	14.22696
PNI_12_35	C. Republic	Rudna	village	50.03559	14.22673
PNI_12_36	C. Republic	Rudna	village	50.03556	14.22665
PNI_12_37	C. Republic	Rudna	village	50.0348	14.22348
PNI_12_38	C. Republic	Kozolupy	village	49.97947	14.18930
PNI_12_39	C. Republic	Unhost	little village	50.0824	14.11584

PNI_12_40	C. Republic	Unhost	little village	50.08237	14.11592
PNI_12_41	C. Republic	Hostivice	village	50.0767	14.24040
PNI_12_42	C. Republic	Hostivice	village	50.07696	14.24078
PNI_12_43	C. Republic	Hostivice	village	50.07739	14.24153
PNI_12_44	C. Republic	Hostivice	village	50.07746	14.24157
PNI_12_45	C. Republic	Hostivice	village	50.07754	14.24179
PNI_12_46	C. Republic	Prague West	city	50.07778	14.30833
PNI_12_47	C. Republic	Prague West	city	50.07778	14.30838
PNI_12_48	C. Republic	Prague West	city	50.07887	14.30817
PNI_12_49	C. Republic	Prague West	city	50.07883	14.30805
PNI_12_50	C. Republic	Prague West	city	50.08705	14.31545
PNI_12_51	C. Republic	Prague West	city	50.08719	14.31551
PNI_12_52	C. Republic	Prague West	city	50.08714	14.31542
PNI_12_53	C. Republic	Prague West	city	50.08715	14.31551
PNI_13_01	Poland	Lipno	village	52.85319	19.16640
PNI_13_02	Poland	Lipno	village	52.85502	19.16340
PNI_13_03	Poland	Lipno	village	52.85475	19.16402
PNI_13_04	Poland	Lipno	village	52.85465	19.16420
PNI_13_05	Poland	Lipno	village	52.85451	19.16444
PNI_13_06	Poland	Kikol	little village	52.90686	19.11993
PNI_13_07	Poland	Kikol	little village	52.90684	19.11987
PNI_13_08	Poland	Kikol	little village	52.90641	19.12048
PNI_13_09	Poland	Kikol	little village	52.90641	19.12046

PNI_13_10	Poland	Lipno	village	52.83156	19.18428
PNI_13_11	Poland	Czarne	meadow	52.79536	19.27204
PNI_13_12	Poland	Fabianci	rural area	52.72771	19.11909
PNI_13_13	Poland	Fabianci	rural area	52.72771	19.11937
PNI_13_14	Poland	Fabianci	rural area	52.72788	19.11871
PNI_13_15	Poland	Fabianci	rural area	52.72787	19.11867
PNI_13_16	Poland	Fabianci	rural area	52.72751	19.11834
PNI_13_17	Poland	Wloclawek	village	52.67197	19.06976
PNI_13_18	Poland	Wloclawek	village	52.67161	19.07131
PNI_13_19	Poland	Wloclawek	village	52.65331	19.11369
PNI_13_20	Poland	Wloclawek	village	52.65323	19.11364
PNI_13_21	Poland	Wloclawek	village	52.65325	19.11367
PNI_13_22	Poland	Wloclawek	village	52.65328	19.11349
PNI_13_23	Poland	Wloclawek	village	52.65317	19.11320
PNI_13_24	Poland	Wloclawek	village	52.65521	19.08436
PNI_13_25	Poland	Wloclawek	village	52.65525	19.08451
PNI_13_26	Poland	Wloclawek	village	52.65526	19.08445
PNI_13_27	Poland	Wloclawek	village	52.65525	19.08467
PNI_14_01	Lithuania	Kiemeliai	meadow	56.07037	24.44008
PNI_14_02	Lithuania	Kiemeliai	meadow	56.0704	24.43994
PNI_14_03	Lithuania	Kiemeliai	meadow	56.07041	24.43986
PNI_14_04	Lithuania	Kiemeliai	meadow	56.07046	24.43998
PNI_14_05	Lithuania	Pasvalys	meadow	56.07036	24.41172

PNI_14_06	Lithuania	Pasvalys	village	56.07034	24.41170
PNI_14_07	Lithuania	Pasvalys	village	56.06986	24.39590
PNI_14_08	Lithuania	Pasvalys	village	56.06986	24.39594
PNI_14_09	Lithuania	Pasvalys	village	56.06987	24.39614
PNI_14_10	Lithuania	Pasvalys	village	56.06989	24.39621
PNI_14_11	Lithuania	Geguzine	meadow	55.84926	24.34988
PNI_14_12	Lithuania	Geguzine	meadow	55.84918	24.34985

10.2 APPENDIX II

CREA-FL Sample list: *Populus nigra var. italica* from different parts of the world		
Clone name/number	Country of origin	Location of collecting site
CREA FL N110	ITA	FONTEVIVO
CREA FL N153	ITA	CASTELNUOVO NE' MONTI
CREA FL N159	ITA	MARZABOTTO
CREA FL N180	ITA	SENIGALLIA (CESANO)
CREA FL N314	ITA	GIULIANOVA
CREA FL N355	ESP	-
CREA FL N360	ITA	GONI
CREA FL N370	ITA	FOCE DEL GARIGLIANO
CREA FL N371	ITA	FOCE DEL GARIGLIANO
CREA FL N372	IRN	LAHIDIAN
CREA FL N376	ITA	CANTALUPO DEL SANNIO
CREA FL N400	ITA	LAGO DI PIEDILUCO
CREA FL N401	PRT	-
CREA FL N410	ITA	GUALDO TADINO
CREA FL N413	ITA	FIUMINATA
CREA FL N414	ITA	-
CREA FL N415	ITA	GUALDO TADINO
CREA FL N417	ITA	-
CREA FL N424	ITA	-
CREA FL N425	USA	-
CREA FL N428	ITA	VALPELLINE
CREA FL N429	ITA	S. GIORGIO MONFERRATO
CREA FL N434	ITA	LEGNAGO
CREA FL N438	GBR	-
CREA FL N446	TUR	-
CREA FL N451	TUR	-
CREA FL N452	TUR	-
CREA FL N488	FRA	AIN
CREA FL N501	FRA	ARDÈCHE
CREA FL N520	FRA	GIRONDE
CREA FL N522	FRA	GIRONDE
CREA FL N524	FRA	GIRONDE
CREA FL N633	BGR	-
CREA FL N641	ITA	LUCCA
CREA FL N644	BGR	-
CREA FL N646	BGR	-
CREA FL N917	ITA	-
CREA FL N918	ITA	-
CREA FL N919	ITA	-
CREA FL N920	ITA	-
CREA FL N921	ITA	-
CREA FL N922	ITA	-
CREA FL N923	ITA	-

CREA FL N924	ITA	-
CREA FL N925	ITA	-
CREA FL N926	ITA	-
CREA FL N927	ITA	-
CREA FL N928	ITA	-
CREA FL N929	ITA	-
CREA FL N930	ITA	-
CREA FL N931	ITA	-
CREA FL N932	ITA	-
CREA FL N933	ITA	-
CREA FL N934	ITA	-
CREA FL N935	ITA	-
CREA FL N936	ITA	-
CREA FL N937	ITA	-
CREA FL N938	ITA	-
CREA FL N952	TUR	-

10.3 APPENDIX III

Group A: Bonafide Lombardy Poplar	N360
	N370
	N371
	N372
	N376
	N410
	N415
	N425
	N428
	N429
	N438
	N488
	N520
	N522
	N524
	N633
	N917
	N918
	N919
	N921
	N922
	N923
	N924
	N925
	N926
	N927
	N928
	N929
	N931
	N935
	N937
	N952
	PNICOMMONGARDEN1--9-02--A8
	PNICOMMONGARDEN1--BG01--A1
	PNICOMMONGARDEN1--BG02--B1
	PNICOMMONGARDEN1--BG03--C1
	PNICOMMONGARDEN1--2-01--D1
	PNICOMMONGARDEN1--2-02--E1
	PNICOMMONGARDEN1--2-04--F1
	PNICOMMONGARDEN1--2-05--G1
	PNICOMMONGARDEN1--2-06--H1

PNICOMMONGARDEN1--2-07--A2
PNICOMMONGARDEN1--2-08--B2
PNICOMMONGARDEN1--3-01--C2
PNICOMMONGARDEN1--3-02--D2
PNICOMMONGARDEN1--3-06--F2
PNICOMMONGARDEN1--3-08--G2
PNICOMMONGARDEN1--3-09--H2
PNICOMMONGARDEN1--3-10--A3
PNICOMMONGARDEN1--3-12--B3
PNICOMMONGARDEN1--3-13--C3
PNICOMMONGARDEN1--3-14--D3
PNICOMMONGARDEN1--3-16--E3
PNICOMMONGARDEN1--4-02--F3
PNICOMMONGARDEN1--4-04--G3
PNICOMMONGARDEN1--4-05--H3
PNICOMMONGARDEN1--4-06--A4
PNICOMMONGARDEN1--4-07--B4
PNICOMMONGARDEN1--4-08--C4
PNICOMMONGARDEN1--4-09--D4
PNICOMMONGARDEN1--4-10--E4
PNICOMMONGARDEN1--4-11--F4
PNICOMMONGARDEN1--4-12--G4
PNICOMMONGARDEN1--5-03--H4
PNICOMMONGARDEN1--5-05--B5
PNICOMMONGARDEN1--5-06--C5
PNICOMMONGARDEN1--5-07--D5
PNICOMMONGARDEN1--5-08--E5
PNICOMMONGARDEN1--5-10--F5
PNICOMMONGARDEN1--5-11--G5
PNICOMMONGARDEN1--5-12--H5
PNICOMMONGARDEN1--6-01--A6
PNICOMMONGARDEN1--6-02--B6
PNICOMMONGARDEN1--6-03--C6
PNICOMMONGARDEN1--6-05--D6
PNICOMMONGARDEN1--6-06--E6
PNICOMMONGARDEN1--6-07--F6
PNICOMMONGARDEN1--6-08--G6
PNICOMMONGARDEN1--6-09--H6
PNICOMMONGARDEN1--7-02--A7
PNICOMMONGARDEN1--7-06--B7
PNICOMMONGARDEN1--7-08--C7
PNICOMMONGARDEN1--7-09--D7
PNICOMMONGARDEN1--7-10--E7

PNICOMMONGARDEN1--7-12--F7
PNICOMMONGARDEN1--7-13--G7
PNICOMMONGARDEN1--9-01--H7
PNICOMMONGARDEN1--9-04--B8
PNICOMMONGARDEN1--9-12--D8
PNICOMMONGARDEN1--9-13--E8
PNICOMMONGARDEN1--9-14--F8
PNICOMMONGARDEN1--10-06--D9
PNICOMMONGARDEN1--10-07--E9
PNICOMMONGARDEN1--10-08--F9
PNICOMMONGARDEN1--10-09--G9
PNICOMMONGARDEN1--10-10--H9
PNICOMMONGARDEN1--11-01--A10
PNICOMMONGARDEN1--11-02--B10
PNICOMMONGARDEN1--11-03--C10
PNICOMMONGARDEN1--11-04--D10
PNICOMMONGARDEN1--11-05--E10
PNICOMMONGARDEN1--11-07--F10
PNICOMMONGARDEN1--11-08--G10
PNICOMMONGARDEN1--11-09--H10
PNICOMMONGARDEN1--11-10--A11
PNICOMMONGARDEN1--12-01--B11
PNICOMMONGARDEN1--12-02--C11
PNICOMMONGARDEN1--12-03--D11
PNICOMMONGARDEN1--12-04--E11
PNICOMMONGARDEN1--12-05--F11
PNICOMMONGARDEN1--12-06--G11
PNICOMMONGARDEN1--12-09--H11
PNICOMMONGARDEN1--12-10--A12
PNICOMMONGARDEN1--12-11--B12
PNICOMMONGARDEN1--12-13--C12
PNICOMMONGARDEN1--13-01--D12
PNICOMMONGARDEN1--13-02--E12
PNICOMMONGARDEN1--13-03--F12
PNICOMMONGARDEN1--13-04--G12
PNICOMMONGARDEN1--13-05--H12
PNICOMMONGARDEN2-2-09-A1
PNICOMMONGARDEN2-2-10-B1
PNICOMMONGARDEN2-2-11-C1
PNICOMMONGARDEN2-3-17-E1
PNICOMMONGARDEN2-3-19-F1
PNICOMMONGARDEN2-3-20-G1
PNICOMMONGARDEN2-3-21-H1

PNICOMMONGARDEN2-3-22-A2
PNICOMMONGARDEN2-3-23-B2
PNICOMMONGARDEN2-3-24-C2
PNICOMMONGARDEN2-3-25-D2
PNICOMMONGARDEN2-3-26-E2
PNICOMMONGARDEN2-3-28-F2
PNICOMMONGARDEN2-3-29-G2
PNICOMMONGARDEN2-3-30-H2
PNICOMMONGARDEN2-3-31-A3
PNICOMMONGARDEN2-3-32-B3
PNICOMMONGARDEN2-3-33-C3
PNICOMMONGARDEN2-4-13-D3
PNICOMMONGARDEN2-4-14-E3
PNICOMMONGARDEN2-4-15-F3
PNICOMMONGARDEN2-4-16-G3
PNICOMMONGARDEN2-4-17-H3
PNICOMMONGARDEN2-4-19-A4
PNICOMMONGARDEN2-4-20-B4
PNICOMMONGARDEN2-4-21-C4
PNICOMMONGARDEN2-4-22-D4
PNICOMMONGARDEN2-4-23-E4
PNICOMMONGARDEN2-5-13-F4
PNICOMMONGARDEN2-5-14-G4
PNICOMMONGARDEN2-5-15-H4
PNICOMMONGARDEN2-5-16-A5
PNICOMMONGARDEN2-5-17-B5
PNICOMMONGARDEN2-5-18-C5
PNICOMMONGARDEN2-5-20-D5
PNICOMMONGARDEN2-5-21-E5
PNICOMMONGARDEN2-5-22-F5
PNICOMMONGARDEN2-6-10-G5
PNICOMMONGARDEN2-6-13-H5
PNICOMMONGARDEN2-6-14-A6
PNICOMMONGARDEN2-6-15-B6
PNICOMMONGARDEN2-6-17-C6
PNICOMMONGARDEN2-6-20-D6
PNICOMMONGARDEN2-6-22-E6
PNICOMMONGARDEN2-6-23-F6
PNICOMMONGARDEN2-7-14-G6
PNICOMMONGARDEN2-7-15-H6
PNICOMMONGARDEN2-7-17-A7
PNICOMMONGARDEN2-7-19-B7
PNICOMMONGARDEN2-7-20-C7

PNICOMMONGARDEN2-7-22-D7
PNICOMMONGARDEN2-7-23-E7
PNICOMMONGARDEN2-9-18-F7
PNICOMMONGARDEN2-9-19-G7
PNICOMMONGARDEN2-9-21-H7
PNICOMMONGARDEN2-9-22-A8
PNICOMMONGARDEN2-9-23-B8
PNICOMMONGARDEN2-9-26-C8
PNICOMMONGARDEN2-9-32-D8
PNICOMMONGARDEN2-11-12-G9
PNICOMMONGARDEN2-11-13-H9
PNICOMMONGARDEN2-11-14-A10
PNICOMMONGARDEN2-11-15-B10
PNICOMMONGARDEN2-11-16-C10
PNICOMMONGARDEN2-11-17-D10
PNICOMMONGARDEN2-11-18-E10
PNICOMMONGARDEN2-11-19-F10
PNICOMMONGARDEN2-12-14-H10
PNICOMMONGARDEN2-12-15-A11
PNICOMMONGARDEN2-12-16-B11
PNICOMMONGARDEN2-12-17-C11
PNICOMMONGARDEN2-12-18-D11
PNICOMMONGARDEN2-12-19-E11
PNICOMMONGARDEN2-12-20-F11
PNICOMMONGARDEN2-12-21-G11
PNICOMMONGARDEN2-12-22-H11
PNICOMMONGARDEN2-12-23-A12
PNICOMMONGARDEN2-13-06-B12
PNICOMMONGARDEN2-13-07-C12
PNICOMMONGARDEN2-13-08-D12
PNICOMMONGARDEN2-13-09-E12
PNICOMMONGARDEN2-13-12-F12
PNICOMMONGARDEN2-14-01-G12
PNICOMMONGARDEN2-14-02-H12
PNICOMMONGARDEN3-2-15-A1
PNICOMMONGARDEN3-2-16-B1
PNICOMMONGARDEN3-2-17-C1
PNICOMMONGARDEN3-2-19-D1
PNICOMMONGARDEN3-3-34-E1
PNICOMMONGARDEN3-3-35-F1
PNICOMMONGARDEN3-3-36-G1
PNICOMMONGARDEN3-3-37-H1
PNICOMMONGARDEN3-3-38-A2

PNICOMMONGARDEN3-3-43-B2
PNICOMMONGARDEN3-3-44-C2
PNICOMMONGARDEN3-3-45-D2
PNICOMMONGARDEN3-3-46-E2
PNICOMMONGARDEN3-4-24-F2
PNICOMMONGARDEN3-4-25-G2
PNICOMMONGARDEN3-4-26-H2
PNICOMMONGARDEN3-4-29-A3
PNICOMMONGARDEN3-4-30-B3
PNICOMMONGARDEN3-4-32-C3
PNICOMMONGARDEN3-4-33-D3
PNICOMMONGARDEN3-4-35-E3
PNICOMMONGARDEN3-4-36-F3
PNICOMMONGARDEN3-4-37-G3
PNICOMMONGARDEN3-4-38-H3
PNICOMMONGARDEN3-4-39-A4
PNICOMMONGARDEN3-4-40-B4
PNICOMMONGARDEN3-4-41-C4
PNICOMMONGARDEN3-4-42-D4
PNICOMMONGARDEN3-4-43-E4
PNICOMMONGARDEN3-5-23-F4
PNICOMMONGARDEN3-5-24-G4
PNICOMMONGARDEN3-5-26-H4
PNICOMMONGARDEN3-5-28-A5
PNICOMMONGARDEN3-5-29-B5
PNICOMMONGARDEN3-5-30-C5
PNICOMMONGARDEN3-5-31-D5
PNICOMMONGARDEN3-5-33-E5
PNICOMMONGARDEN3-5-34-F5
PNICOMMONGARDEN3-6-24-G5
PNICOMMONGARDEN3-6-27-H5
PNICOMMONGARDEN3-6-28-A6
PNICOMMONGARDEN3-6-29-B6
PNICOMMONGARDEN3-6-31-C6
PNICOMMONGARDEN3-6-32-D6
PNICOMMONGARDEN3-6-33-E6
PNICOMMONGARDEN3-6-34-F6
PNICOMMONGARDEN3-7-24-G6
PNICOMMONGARDEN3-7-26-H6
PNICOMMONGARDEN3-7-30-A7
PNICOMMONGARDEN3-7-33-B7
PNICOMMONGARDEN3-7-34-C7
PNICOMMONGARDEN3-7-35-D7

PNICOMMONGARDEN3-9-33-E7
PNICOMMONGARDEN3-9-35-F7
PNICOMMONGARDEN3-9-36-G7
PNICOMMONGARDEN3-9-37-H7
PNICOMMONGARDEN3-9-38-A8
PNICOMMONGARDEN3-9-39-B8
PNICOMMONGARDEN3-9-41-C8
PNICOMMONGARDEN3-10-21-D8
PNICOMMONGARDEN3-10-23-E8
PNICOMMONGARDEN3-10-24-F8
PNICOMMONGARDEN3-10-25-G8
PNICOMMONGARDEN3-10-27-H8
PNICOMMONGARDEN3-10-28-A9
PNICOMMONGARDEN3-10-30-C9
PNICOMMONGARDEN3-10-32-D9
PNICOMMONGARDEN3-10-33-E9
PNICOMMONGARDEN3-11-25-H9
PNICOMMONGARDEN3-11-26-A10
PNICOMMONGARDEN3-11-27-B10
PNICOMMONGARDEN3-11-28-C10
PNICOMMONGARDEN3-11-29-D10
PNICOMMONGARDEN3-11-30-E10
PNICOMMONGARDEN3-11-31-F10
PNICOMMONGARDEN3-12-24-G10
PNICOMMONGARDEN3-12-25-H10
PNICOMMONGARDEN3-12-26-A11
PNICOMMONGARDEN3-12-27-B11
PNICOMMONGARDEN3-12-28-C11
PNICOMMONGARDEN3-12-29-D11
PNICOMMONGARDEN3-12-31-E11
PNICOMMONGARDEN3-12-32-F11
PNICOMMONGARDEN3-12-33-G11
PNICOMMONGARDEN3-12-34-H11
PNICOMMONGARDEN3-13-13-A12
PNICOMMONGARDEN3-13-14-B12
PNICOMMONGARDEN3-13-15-C12
PNICOMMONGARDEN3-13-18-D12
PNICOMMONGARDEN3-13-19-E12
PNICOMMONGARDEN3-14-03-F12
PNICOMMONGARDEN3-14-04-G12
PNICOMMONGARDEN3-14-06-H12
PNICOMMONGARDEN4-2-22-A1
PNICOMMONGARDEN4-2-23-B1

PNICOMMONGARDEN4-2-24-C1
PNICOMMONGARDEN4-3-47-D1
PNICOMMONGARDEN4-3-48-E1
PNICOMMONGARDEN4-3-50-F1
PNICOMMONGARDEN4-3-51-G1
PNICOMMONGARDEN4-3-53-H1
PNICOMMONGARDEN4-3-57-A2
PNICOMMONGARDEN4-3-58-B2
PNICOMMONGARDEN4-3-59-C2
PNICOMMONGARDEN4-4-44-D2
PNICOMMONGARDEN4-4-45-E2
PNICOMMONGARDEN4-4-46-F2
PNICOMMONGARDEN4-4-47-G2
PNICOMMONGARDEN4-4-49-H2
PNICOMMONGARDEN4-4-50-A3
PNICOMMONGARDEN4-4-51-B3
PNICOMMONGARDEN4-4-52-C3
PNICOMMONGARDEN4-4-55-D3
PNICOMMONGARDEN4-4-56-E3
PNICOMMONGARDEN4-5-35-F3
PNICOMMONGARDEN4-5-36-G3
PNICOMMONGARDEN4-5-38-H3
PNICOMMONGARDEN4-5-39-A4
PNICOMMONGARDEN4-5-40-B4
PNICOMMONGARDEN4-5-41-C4
PNICOMMONGARDEN4-5-44-D4
PNICOMMONGARDEN4-5-50-E4
PNICOMMONGARDEN4-5-51-F4
PNICOMMONGARDEN4-5-52-G4
PNICOMMONGARDEN4-5-53-H4
PNICOMMONGARDEN4-5-54-A5
PNICOMMONGARDEN4-5-56-B5
PNICOMMONGARDEN4-6-35-C5
PNICOMMONGARDEN4-6-38-D5
PNICOMMONGARDEN4-6-39-E5
PNICOMMONGARDEN4-6-41-F5
PNICOMMONGARDEN4-6-44-G5
PNICOMMONGARDEN4-6-47-H5
PNICOMMONGARDEN4-6-48-A6
PNICOMMONGARDEN4-6-51-B6
PNICOMMONGARDEN4-6-52-C6
PNICOMMONGARDEN4-7-36-D6
PNICOMMONGARDEN4-7-37-E6

PNICOMMONGARDEN4-7-38-F6
PNICOMMONGARDEN4-7-39-G6
PNICOMMONGARDEN4-7-40-H6
PNICOMMONGARDEN4-7-41-A7
PNICOMMONGARDEN4-7-43-B7
PNICOMMONGARDEN4-9-44-C7
PNICOMMONGARDEN4-9-45-D7
PNICOMMONGARDEN4-9-47-E7
PNICOMMONGARDEN4-9-49-F7
PNICOMMONGARDEN4-9-50-G7
PNICOMMONGARDEN4-10-34-B8
PNICOMMONGARDEN4-10-35-C8
PNICOMMONGARDEN4-10-36-D8
PNICOMMONGARDEN4-10-37-E8
PNICOMMONGARDEN4-10-38-F8
PNICOMMONGARDEN4-10-39-G8
PNICOMMONGARDEN4-10-40-H8
PNICOMMONGARDEN4-10-41-A9
PNICOMMONGARDEN4-10-42-B9
PNICOMMONGARDEN4-10-43-C9
PNICOMMONGARDEN4-11-32-D9
PNICOMMONGARDEN4-11-33-E9
PNICOMMONGARDEN4-11-34-F9
PNICOMMONGARDEN4-11-35-G9
PNICOMMONGARDEN4-11-37-H9
PNICOMMONGARDEN4-11-38-A10
PNICOMMONGARDEN4-11-39-B10
PNICOMMONGARDEN4-11-40-C10
PNICOMMONGARDEN4-11-41-D10
PNICOMMONGARDEN4-12-35-E10
PNICOMMONGARDEN4-12-36-F10
PNICOMMONGARDEN4-12-37-G10
PNICOMMONGARDEN4-12-38-H10
PNICOMMONGARDEN4-12-39-A11
PNICOMMONGARDEN4-12-40-B11
PNICOMMONGARDEN4-12-41-C11
PNICOMMONGARDEN4-12-42-D11
PNICOMMONGARDEN4-12-43-E11
PNICOMMONGARDEN4-12-44-F11
PNICOMMONGARDEN4-13-20-G11
PNICOMMONGARDEN4-13-21-H11
PNICOMMONGARDEN4-13-22-A12
PNICOMMONGARDEN4-13-24-C12

PNICOMMONGARDEN4-14-07-D12
PNICOMMONGARDEN4-14-08-E12
PNICOMMONGARDEN4-14-09-F12
PNICOMMONGARDEN4-14-10-G12
PNICOMMONGARDEN4-14-11-H12
PNICOMMONGARDEN5-3-60-A1
PNICOMMONGARDEN5-4-34-B1
PNICOMMONGARDEN5-5-09-C1
PNICOMMONGARDEN5-5-37-D1
PNICOMMONGARDEN5-5-46-E1
PNICOMMONGARDEN5-6-25-F1
PNICOMMONGARDEN5-6-53-G1
PNICOMMONGARDEN5-6-54-H1
PNICOMMONGARDEN5-6-56-B2
PNICOMMONGARDEN5-7-28-C2
PNICOMMONGARDEN5-7-50-D2
PNICOMMONGARDEN5-7-51-E2
PNICOMMONGARDEN5-7-52-F2
PNICOMMONGARDEN5-7-53-G2
PNICOMMONGARDEN5-9-11-H2
PNICOMMONGARDEN5-10-44-F3
PNICOMMONGARDEN5-10-45-G3
PNICOMMONGARDEN5-10-46-H3
PNICOMMONGARDEN5-10-47-A4
PNICOMMONGARDEN5-10-49-B4
PNICOMMONGARDEN5-10-50-C4
PNICOMMONGARDEN5-10-53-F4
PNICOMMONGARDEN5-11-42-G4
PNICOMMONGARDEN5-11-44-H4
PNICOMMONGARDEN5-11-45-A5
PNICOMMONGARDEN5-11-46-B5
PNICOMMONGARDEN5-11-47-C5
PNICOMMONGARDEN5-11-48-D5
PNICOMMONGARDEN5-11-53-F5
PNICOMMONGARDEN5-12-45-G5
PNICOMMONGARDEN5-12-46-H5
PNICOMMONGARDEN5-12-47-A6
PNICOMMONGARDEN5-12-48-B6
PNICOMMONGARDEN5-12-49-C6
PNICOMMONGARDEN5-12-51-E6
PNICOMMONGARDEN5-12-52-F6
PNICOMMONGARDEN5-13-11-H6
PNICOMMONGARDEN5-13-25-A7

	PNICOMMONGARDEN5-13-26-B7
	PNICOMMONGARDEN5-14-12-D7

Group B: Italian Columnar Populus nigra italica, coming from second introduction/local mutation	PNICOMMONGARDEN1--10-01--G8
	PNICOMMONGARDEN1--10-02--H8
	PNICOMMONGARDEN1--10-03--A9
	PNICOMMONGARDEN1--10-04--B9
	PNICOMMONGARDEN1--10-05--C9
	PNICOMMONGARDEN2-10-11-E8
	PNICOMMONGARDEN2-10-12-F8
	PNICOMMONGARDEN2-10-13-G8
	PNICOMMONGARDEN2-10-14-H8
	PNICOMMONGARDEN2-10-15-A9
	PNICOMMONGARDEN2-10-16-B9
	PNICOMMONGARDEN2-10-17-C9
	PNICOMMONGARDEN2-10-18-D9
	PNICOMMONGARDEN2-10-19-E9
	PNICOMMONGARDEN2-10-20-F9
	PNICOMMONGARDEN2-11-22-G10
	PNICOMMONGARDEN3-11-23-F9
	PNICOMMONGARDEN3-11-24-G9
	PNICOMMONGARDEN5-10-51-D4
	PNICOMMONGARDEN5-10-52-E4
	PNICOMMONGARDEN5-11-52-E5
	N180
	N314
	N400
	N413
	N414
	N417
	N641
	N920
	N930
	N933
	N934
	N938

Parent Other introductions or local mutants in EU	N153
Offspring San giorgio X European P. nigra population	N424
Offspring San giorgio X European P. nigra population	N434
Offspring San giorgio X European P. nigra population	N644
Offspring San giorgio X European P. nigra population	N645
Offspring San giorgio X European P. nigra population	N646
Offspring San giorgio X European P. nigra population	N932
Offspring San giorgio X European P. nigra population	PNICOMMONGARDEN1--3-04--E2
Female parent of San giorgio	PNICOMMONGARDEN2-2-12-D1
Offspring San giorgio X italian P. nigra population	PNICOMMONGARDEN3-10-29-B9
SAN GIORGIO	PNICOMMONGARDEN4-13-23-B12
Low Coverage	PNICOMMONGARDEN5-12-50-D6
Low Coverage	PNICOMMONGARDEN5-12-53-G6
Low Coverage	PNICOMMONGARDEN1--5-04--A5
Low Coverage	PNICOMMONGARDEN1--9-10--C8
Low Coverage	PNICOMMONGARDEN5-6-55-A2
P. nigra hybrid with other poplar species	PNICOMMONGARDEN4-9-52-H7
P. nigra hybrid with other poplar species	PNICOMMONGARDEN4-9-53-A8
P. nigra hybrid with other poplar species	PNICOMMONGARDEN5-9-54-A3
P. nigra hybrid with other poplar species	PNICOMMONGARDEN5-9-57-B3
P. nigra hybrid with other poplar species	PNICOMMONGARDEN5-9-58-C3
P. nigra hybrid with other poplar species	PNICOMMONGARDEN5-9-59-D3
P. nigra hybrid with other poplar species	PNICOMMONGARDEN5-9-62-E3

10.4 APPENDIX IV

Bio Sample	Organism	Center_Name	geo_loc_name
SAMN01885652	Populus euphratica	BGI	
SAMN03196385	Populus tremula x Populus alba	UNIVERSITY OF GEORGIA	USA:Athens,GA
SAMN04299929	Populus nigra	CLINVAR-IGA TECHNOLOGY SERVICES	Italy: Ticino river
SAMN04534693	Populus simonii	NANJING FORESTRY UNIVERSITY	China:Jiangsu
SAMN06011208	Populus pruinosa	SICHUAN UNIVERSITY	China:Xin Jiang
SAMN07334578	Populus alba var. pyramidalis	LANZHOU UNIVERSITY	China: Lanzhou University, Gansu
SAMN08013139	Populus deltoides x Populus simonii	NANJING FORESTRY UNIVERSITY	China: Jiangsu
SAMN09228080	Populus ilicifolia	SICHUAN UNIVERSITY	Kenya
SAMN09072051	Populus trichocarpa	JGI	missing
SAMN10068037	Populus tomentosa	CHINESE ACADEMY OF FORESTRY	China: Hebei Province
SAMN11620760	Populus deltoides	CSIRO	not applicable
SAMN11775866	Populus tremula	THUENEN INSTITUTE OF FOREST GENETICS	Germany
SAMN11775864	Populus alba	THUENEN INSTITUTE OF FOREST GENETICS	Germany
SAMN11775868	Populus tremuloides	THUENEN INSTITUTE OF FOREST GENETICS	Germany
SAMN10818489	Populus davidiana	CHINESE ACADEMY OF FORESTRY	China:Jilin
SAMN12347741	Populus alba x Populus glandulosa	CHINESE ACADEMY OF FORESTRY	China: Beijing
SAMN14379744	Populus rotundifolia	UNIVERSITY OF VIENNA	China: LSichuan
SAMN14379755	Populus adenopoda	UNIVERSITY OF VIENNA	China: Hunan
SAMN14379758	Populus qiongdaoensis	UNIVERSITY OF VIENNA	China: Hainan
SAMN15030613	Populus glandulosa	CHINESE ACADEMY OF FORESTRY	South Korea: Suwon
SAMN14485777	Populus balsamifera	UNIVERSITY OF COPENHAGEN	Denmark
SAMN15555026	Populus koreana	WEST VIRGINIA UNIVERSITY	China: Kuandian County
SAMN17849155	Populus angustifolia	GUANGXI MEDICAL UNIVERSITY	China: Chengdu

SAMN17005496	Populus szechuanica	SOUTHWEST MEDICAL UNIVERSITY	China:SiChuan
SAMN17808122	Populus lasiocarpa	LUOHE MEDICAL COLLEGE	China: Chengdu
SAMN01924667	Populus x canadensis	BIOLABS, INSTITUTE OF LIFE SCIENCE, SCUOLA SUPERIO	
SAMN15454199	Populus x sibirica	EMB, RUSSIAN ACADEMY OF SCIENCES	Russia: Moscow
SAMN14485780	Populus x berolinensis	UNIVERSITY OF COPENHAGEN	Denmark

11 ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my supervisor, Prof. Michele Morgante, for providing me with the opportunity to flourish in the field of genomics and for guiding me throughout this journey. Dear Professor, your words, "If you do not understand why you are doing something, you are wasting your time doing it," will stay with me forever.

I am also deeply thankful to my co-supervisors, Dr. Gabriele De Gaspero and Dr. Davide Scaglione, for their constant support and assistance in navigating the complexities of my research and helping me comprehend the procedures.

I extend my sincere appreciation to Dr. Gabriele Magris, whose presence and support were invaluable. Gabbo, you were always there for us, whether it was for academic endeavours or simply enjoying wine tasting and food exploration. Your assistance in developing and executing complicated scripts and file manipulations was truly appreciated.

I would like to acknowledge the EpiDiverse Project and its members for their help and support—special thanks to the EU Horizon 2020 program for providing financial support for my research. I am grateful to the entire team, especially the Populus Group (Barbara and Cristian), for providing valuable samples. I also acknowledge the Consiglio per la ricerca in agricoltura e l'analisi dell'economia agrarian-FL, Casale Monferrato, for providing the Populus nigra var. italica cuttings from global locations. To the EpiDiverse team, thank you for the enjoyable times we shared during summer schools and meetups.

I extend my appreciation to Dr. David Langenberger, Dr. Adam Nun at ecSeq, and Dr. Marie Mirouze, Dr. Panpan Zhang at the University of Perpignan for their support and the opportunity to engage in international secondments.

A special thank you to everyone at the Institute of Applied Genomics and IGA Technology Services at Udine IT for their administrative and technical assistance with my research.

To my beloved daughter, Atishi Dubay Pandey, you have been my inspiration to persevere. Your presence in my life opened a new chapter, and I am forever grateful for your love and encouragement.

My deepest gratitude goes to my family and friends, whose unwavering emotional support, inspiration, and determination have been instrumental throughout my academic journey. I am eternally thankful for their love and patience.

To my spouse, Mr. Vishwa Deep Pandey, thank you for supporting my PhD dream despite the distance between us. To my brother, Dr. Mayank Sharma, your guidance during challenging times has been invaluable.

I am grateful to my parents, Mr. Y.K. Sharma and Mrs. Madhu Sharma, for being my pillars of strength. A heartfelt thank you to my In-Laws for their unwavering support and encouragement, even in an orthodox society.

I cannot go without expressing my appreciation to Fabio for his constant moral support as a dear friend for life. To Ariana, Emanuele, Simone, M. Vidotto, and Paloma, I am thankful for your presence in my life during my time in Udine, including the challenges of lockdown and the COVID pandemic.

Lastly, my committee members deserve my gratitude for their valuable time, constructive critique, and insightful comments on my research, which greatly contributed to the development of my ideas and arguments.