Inland Norway University of Applied Sciences

Faculty of Applied Ecology, Agriculture and Biotechnology

Prativa Paudel

Master's Thesis

Genetic and epigenetic analyses of Scots pine of divergent origin, sowed Swedish seed vs. wind-blown Norwegian seed in two fire-cleared and two logged forests

Master in Applied and Commercial Biotechnology

2024

Consent to lending by University College Library YES ☒ NO ☐

Consent to accessibility in digital archive Brage YES ☒ NO ☐

Acknowledgement

Completion of this thesis has been a challenging yet rewarding journey and was not possible without the guidance, support, and encouragement from many individuals - for each of them I am deeply grateful.

First and foremost, I would like to begin by expressing my sincere appreciation and gratitude to my major supervisor Professor Robert Wilson for his unwavering support, invaluable and patient guidance, encouragement, and insightful feedback, which has been instrumental in shaping this research and pushed me to achieve more than I thought. I am immensely thankful to another member of my thesis committee, Associate Professor Hanne K. Sjølie (leader of the Scots pine growth project funded by the Norwegian "Skogtiltaksfondet"), for the thoughtful suggestions and collaborations, providing a supportive environment for this thesis. I also want to extend my gratitude to Glommen Mjøsen Skog SA – along with Sverre Holm (R&D director), & Marie Kalstad (for the sampling), and Chedley Kastally, for their crucial help in the completion of this thesis. Further, I would like to acknowledge the research environment and opportunities provided by Inland Norway University of Applied Sciences (Campus Hamar) to carry out this thesis.

I owe a deep dept of gratitude to my family for their understanding, unconditional love, and encouragement throughout my academic journey. And a special thanks to my dear husband, Ujjwal for the endless patience, constant support, love, and source of strength throughout this journey.

Abbreviations

SNP	Single nucleotide polymorphism
N v S	Norwegian versus Swedish
T v S	Tall versus Short
GWAS	Genome wide association study
SPEI	Standardized precipitation evapotranspiration index
HTLP	High temperature and low precipitation
LTHP	Low temperature and high precipitation
GO	Gene Ontology
RRBS	Reduce representative bisulfite sequencing
DMRs	Differentially methylated regions
RefFreeDMA	Reference free DNA methylation analysis
BP	Biological process
MF	Molecular function
CC	Cellular component
DI	Dominant interaction
RI	Recessive interaction
DWC	Dominant without covariance
RWC	Recessive without covariance
GWC	Genotypic without covariance
MAF	Minor allele frequency

HWE	Hardy-Weinberg equilibrium
PCA	Principal component analysis
BLAST	Basic local alignment search tool

Table of Contents

Abstract

Scots pine (*Pinus sylvestris* L.) is the most widely distributed pine species across Eurasia and is a vital genetic resource in Norway with a widespread geographic range. It is one of the most suitable and best pioneer species for the afforestation because of its undemanding nature and higher adaptability potential in poor soil conditions. This study investigated population structure, GWAS (Genome wide association studies), and Gene ontology using 50K SNP genotyping data as well as DNA methylation analyses from reduced representation bisulfite sequencing (RRBS) data. Population structure was estimated using fastSTRUCTURE, STRUCTURE 2.3.4 and STRUCTURE HARVESTER, GWAS was performed using Plink v2.0, gene ontology through R-stats and DNA methylation through RefFreeDMA (Reference free DNA methylation analysis) software, further differentially methylated regions functionally characterized via BLAST algorithms using QIAGEN CLC Genomic Workbench 22.0.2. Population structure analysis of 310 individual samples identified the most likely K-value of 8, and genetic diversity within all the cluster showed significant variability; mean FST values indicated genetic differentiation, particularly for clusters 5 and 8. Cluster 4 was the largest with 240 of 310 trees and was therefore presumed to represent trees originating from the sowed Swedish seed. GWAS identified distinct SNP markers associated with plant height (PH) based on different allele and genotype models, with markers in the dominant allele model predominating. Interaction analysis further explained allele-specific associations particularly observed in Swedish trees across the forest plots. Analyses were expanded to investigate influence of seasonal variation in annual growth of plants by correlating SNP markers, growth associated SNPs varied significantly across seasons; dominant allele models showed highest associations during periods of High SPEI. Gene ontology (GO) analysis provided insights into biological importance of identified genes, identifying different molecular functions, biological processes, and cellular components associated with PH. Particularly, a subset of genes showed shared and unique functions across different samples and allele models, emphasizing detailed genetic mechanisms underlying PH and plant growth. RRBS analyses of 29 sampled trees, roughly 50:50 Swedish:Norwegian and 50:50 tall:short, identified 394 differentially methylated fragments between Norwegian and Swedish seed origin, 129 in CpG, 128 in CHG, and 80 in CHH. Norwegian trees showed higher prevalence of methylation in CHH site, while Swedish trees showed higher methylation in CpG. In the tree height comparison, 593 DMRs were identified between tall and short trees: 293 in CpG, 188 in CHG, and 112 in CHH. Individuals in both tree growth categories (tall and short) exhibited high methylated sites in CHH, while fewer methylated sites were observed in CHG and CpG for tall trees and short trees, respectively. Therefore, our detailed analysis enhances the understanding of underlying genetic and epigenetic mechanisms regulating Scots pine growth traits. And the findings of this study provide valuable insights for conservation practices and sustainable forest management as well as breeding approaches directed towards enhancing resilience and productivity in *Pinus sylvestris*.

Keywords: Scots Pine; *Pinus sylvestris*; SNP; population structure; GWAS; gene ontology; DNA methylation

1. Introduction

1.1 Scots Pine (*Pinus Sylvestris* L.)

Pinus sylvestris L. (Scots Pine) is a two-needle pine which belongs to family Pinaceae and the most widely distributed pine species - across Eurasia, North America to Central America, North Africa, and nearby islands. The word "*sylvestris*" came from the Latin meaning "of forests" and its distribution in wide range suggests higher genetic variation and existence of many sub-species and varieties (Durrant et al., 2016). Scots pine can be easily distinguished with its orange-red bark colour, where the upper part of stem has thin, papery, and bright orange or reddish-brown bark, whereas the lower part of stem has thick brown coloured bark that is deeply fissured in young trees and shallow fissured in old trees with thin, pale grey-brown coloured bark. It has a monopodial growth with a termina bud on aerial shoots surrounded by several lateral buds, in which the branches are produced in whorls, where lateral buds between the whorl is absent and the needle-shaped leaves occur in pairs which arises from a short shoot with the leaf-bases surrounded by 5-10 mm long common sheath. Needles are generally slightly twisted and finely serrulate, 3-8 cm long and 1-2 mm wide with grey or blue- green summer colour, dark grey-green or yellow, or reddish green in winter which persist 2-4 years in general to up to 6 years. The needles are xeromorphic with imbedded stomata, a well-developed collenchymatous hypodermis with a waxy layer on thick-walled epidermis, so that water loss is minimized, helping it to adapt to cold and drought conditions (Carlisle & Brown, 1968). Scots pine is a monoecious species in general, but only male or only female flowers in mature trees might occur occasionally, in which wind pollination prevails mainly and the pollen is round or elliptic in shape, size of 44-52 µm with two wings. At the base of current season's growth (new shoot) on older shaded branches with low vigour, clusters of male inflorescences are formed (when buds open in early spring), which are ovoid, 4-8 mm long and yellow or violet-yellow to occasionally pink or violet-red colour pre-anthesis. Whereas the female inflorescence or cone is formed (mid-May to early June) on the more vigorous and well-illuminated shoots i.e., apex of new and strong shoots that are borne at end of current year's growth having seminiferous scales with red or rose-purple shade, 5-6 mm long and pointed tips. First time flowering occurs in the spring or early summer (April - June) when the tree is around 15 years and the female flowers are pollinated after the anthesis in about June and when matured, fertilization occurs the year following pollination i.e., upcoming spring, it forms woody cones of 5-8 cm conic-oblong and in each ovuliferous scale with two integumental ovules winged seed of grey, brown, or yellow colour is formed after fertilization. When the environment is favourable i.e., alternate dry and wet periods (over 9-12 months), the cones open, the ripened winged seeds are released, and dispersed from parent trees. Under favourable conditions, epigeal germination of seeds occurs within 10 days, cotyledons fully develop in about 10-16 days, apical bud advances to form primordial leaves - on whose axils the bifoliate fascicles develop to grow as an adult foliage in about 8-10 months post germination, after which the bifoliate shoots developed from brown scales axils produces the foliage in plants (Carlisle & Brown, 1968; Durrant et al., 2016; Brichta et al., 2023).

Figure 1. Representation of morphology of Scots pine (Pinus sylvestris L.) showing (A) tree, (B) section of branch, (C) female cone, (D) male strobili, (E) needle, and (F) seed (Figure adopted from Brichta et al., 2023).

1.2 Habitat, Ecology, Importance and Usage

Scots pine, essentially a gregarious and light demanding species, doesn't tolerate heavy shading, atmospheric pollution or salty sea wind but can thrive well and colonize in nutrient poor and disturbed sites, drought as well as frost condition, and acidic highland moors. With the rising temperature it's growth is more towards northern region while decreasing in its southern inhabitation and it requires a period of winter chilling after autumn to break it's dormancy and again starts to grow when temperature rises around 5 °C in spring (Durrant et al., 2016). It lives up to 200-300 years and grows on average to 23–27 m in height, and 50–80 cm in diameter with yellowish-white sapwood, and pale brown or reddish heartwood of stem. The crown forms of Scots pine vary, in Scandinavia and higher altitude it is "spruce-like" i.e., narrow cone shaped and finely branched. Being the widely spread species within genus Pinus, it can be found from 8° W in Spain to 141° E of Russia to 70° N in Scandinavia covering 10,000 km longitudinally and representing the distance of 3700 km. It can grow between 200 – 2000 mm of annual precipitation with 300 mm as essential limit, 5-9 °C common mean annual temperature, minimum of 100 frost free is required for its survival and can be found on very poor to rich acid or alkaline soil condition mostly on sandy, sandy loam, and peat soils, but on fertile soils it is outnumbered by another species. It has been successfully exploiting the less crowded ecological niche and reproducing in unstable environment mainly due to its set of traits, such as (a) its low nutrient demand is supported by association with mycorrhiza in low nutrient soils, (b) light demanding nature and rapid growth of youth makes it a pioneer colonizer in open areas, (c) even a single tree can propagate due to its monoecious nature facilitating in-breeding and selfing, and (d) long distance pollination and seed transport is effective with wind, so reproduction is almost guaranteed with annual seed production with

short generation time (Krakau et al., 2013). Scots pine can change the anatomical and morphological structure of the needle based on growing environment which in-turn determines the growth duration and intensity of various organs, and the productivity as well as stability of stands. The morphological changes include the length and volume of the needle i.e., long or short, dense or thinner, wider or narrower, resilient or fragile, while the anatomical changes include the thickness of assimilative & conductive tissues, of folded mesophyll, diameter of vascular bundles and resin canals, number of resin canals, and thickness of external protective tissues of the needle (Galdina & Khazova, 2019). In Europe, monocultures of Scots pine are found in poorest habitat, but in uplands or slightly richer habitats it can be found in mixture with oaks (*Quercus petraea*, *Quercus robur*), Norway spruce (*Picea abies*), silver birch (*Betula pendula*), European larch (*Larix decidua*), European beech (*Fagus sylvatica*), silver fir (*Abies alba*), and other pines (primarily *Pinus uncinate*, *Pinus nigra*) (Brichta et al., 2023; Durrant et al., 2016). It stores food primarily in the form of fat (oils) rather than as starch, and in winter starch is converted to fat, but in summer when the temperature rises fat is converted back to starch. Also, turpentine as a part of resins is produced as a product of certain metabolic processes in Scots pine, commercially tapped, and contains α-pinene, some β-pinene and Δ^3-carene, limonene, terpinolene and di-pentene in small quantities (Carlisle & Brown, 1968). These are the component of essential oils which can also be obtained through hydro-distillation of needles and have high antimicrobial activity (α-pinene, β-pinene, limonene), insect larvicidal potential, and limonene emulsion can be applied to prolong the self-life of food as it avoids foodborne pathogen contamination (Ibáñez & Blázquez, 2019). Also, the different components of essential oils have been used as a starting material for varnishes, paint thinning solvents, printing inks, insecticides, rubbers, and the pleasant aroma of essential oil helps to calm down nervous system and used in the perfumes, air fresheners, soaps, and massage oils. Its wider distribution and ecological plasticity make it one of the important tree species for forestry and wood industry in central Europe and throughout Eurasia, while the density of spring and summer wood varies between 0.412 and 0.541 g cm^{-3}. The distinct and differentiated summer ring is due to high resin content which in turn makes the wood very durable making it very useful material in water and humid conditions as well as in lumber, timber constructions, and furniture, while the low-quality woods can be used for fuel and fibre and bark could be used as an insulating material for buildings (Brichta et al., 2023). Scots pine can be the most suitable and best pioneer species for the afforestation of the reclamation sites, such as post-mining areas and climate change impacted sites because of its undemanding nature and higher adaptability potential in poor soil conditions (Vacek et al., 2021).

1.3 Status and activities on Scots Pine in Norway

The present status and composition of forest is resulted from the interaction of human and forest over long periods of time in past, such as deforestation, induced regeneration, and silviculture practices, so it varies from forests of primeval time. Different factors determined the species composition and their subsequent distribution in Norwegian forest, such as tree species immigration after the Ice Age, successive climatic changes, and human interactions. Likewise, studies pertaining to evidence of ancient DNA from lake sediments and pollen,

coupled with recent DNA samples have supported the hypothesis that Pinus sylvestris survived the ice-free refugia of Scandinavia throughout the course of most recent glaciation and *Pinus sylvestris* was present as early as 22,000 years ago on the north-western coast of Norway. *Pinus sylvestris* is considered as vital in genetic resource category and is found in widespread geographic range of Norway, which occurs in strand where wind acts mainly as an effective pollinator and seed dispersal agent (Fjellstad & Skrøppa, 2020).

In Norway, forest covers 12.1 million hectares of land, and the Norwegian forests can be differentiated into three major forest type categories with their main tree species, namely coniferous evergreen boreal forest (5.8 million hectares - *Picea abies*, *Pinus sylvestris*), broadleaved forest (4.2 million hectares - *Betula pubescens*, *B. pendula*), and mixed forest (2.1 million hectares - *P. abies, P. sylvestris, Betula sp*.). Moreover, the coniferous forest is formed mainly of *Picea abies* (Norway spruce), and *Pinus sylvestris* (Scots pine), which covers 48 % of forest area and 88 % of the annual forest fellings. These forest species dominate the inland region of Norway and are economically the most important, as well as the only species which are managed actively in commercial forestry for wood production. Apart from forests, 2.1 million hectares of other wooded land is present in Norway, and 42.7 % (2728 out of 6396 growing stock 1000 m^3 over bark) of the growing stock on those land is accounted by the *Pinus sylvestris* trees of diameter (≥ 5 cm) at breast height (dbh) equal to 1.3 meter. In Norway, the genetic variability of *Pinus sylvestris* has been evaluated at family level (offspring from one mother tree, with the evaluated samples from provenances - trees from a defined geographic region) whose morphological traits, adaptive and production characters have been assessed, however molecular characterization has not been performed yet. A recent survey on the number of published studies of genetic variability of native Norwegian tree species (from 1954 – 2019) showed that the studies on *Pinus sylvestris* were mostly conducted from the 1950s to early 1990s, however in recent years the studies have been carried out mainly on Norway spruce (Picea *abies*), and Birch (*Betula species*) with more focus on other broadleaved species for their better management. Field tests in forest genetic research is mainly performed by The Norwegian Institute of Bioeconomy Research (NIBIO), The Norwegian University of Life Sciences (NMBU) and The Norwegian Forest Seed Centre and the collection of *Pinus sylvestris* in provenance tests have been carried out from 6 stands and 20 accessions, which are not part of the breeding programmes. As an in-situ conservation effort, genetic characterization of *Pinus sylvestris* along with other species has been performed in five gene conservation units (GCUs) (in already established nature conservation areas) as a part of the Horizon 2020 GenTree project and regular monitoring has been executed since 2020, so that the data can be utilized in the future in connection with genetic monitoring. Forest genetic resources in Norway are conserved in ex-situ conservation stands and in seed storage, as a part of Ex situ conservation efforts, with proper strategy for long term regeneration, however the major focus is on Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*) as important commercial species. The ex-situ conservation stands are maintained with a cooperation between the Norwegian Genetic Resource Centre, The Norwegian Forest Seed Centre and forest owners with an aim to conserve original forest genetic resources through regular forestry as well as regeneration, selection of trees for breeding, and seed collection. For seed storage, seed samples of Scots pine and Norway spruce were deposited in Svalbard Global Seed Vault

(SGSV) in 2015, primarily for back-up storage to conserve and monitor changes in genetic diversity from various stages and generation of breeding populations or seed orchards, with discussion on expanding its use for future monitoring and conservation. Moreover, the Norwegian Forest Seed Center established a biobank with 770 reference seed and DNA samples of spruce and pine from all seed lots available a decade ago for experiments and research, having a representative selection of those seed lots stored at SGSV (Fjellstad & Skrøppa, 2020).

1.3.1 The State of use, genetic improvement and breeding

In production forestry, after the harvest, *Pinus sylvestris* is mainly regenerated naturally executing seed-tree method in which 30-150 seed trees are retained per hectare as a regeneration felling strategy. Primarily, *Pinus sylvestris* planting is based on seeds from the natural forest, but incorporation of genetically improved materials from the Swedish tree improvement program has been of recent interest. The Norwegian Forest Seed Centre (Skogfrøverket) is a responsible organization for tree breeding activities which also provides the seed required for forest nurseries, offers guidance for utilizing materials from national seed orchards and is a responsible body for acquisition, storage, and trading of seeds for forestry sector i.e., seed sales – domestic and international transfer (imported and exported). Norwegian forest nurseries located at different regions produce seedlings as per local requirement, in general, and seedlings of *Pinus sylvestris* are also imported, mainly from Sweden. For the year 2021 and 2022, domestic seed sale and export of native *P. sylvestris* were 37.13 kg & 0 kg, and 33.67 kg & 0.10 kg, respectively, while domestic seed sale and export of exotic *P. sylvestris* were 8.55 kg & 0 kg, and 7.43 kg & 0 kg, respectively. Out of 41.20 kg total seed sales in 2022, 11.33 kg and 29.87 kg were sold for seed plantation (OECD category - 5.83 kg qualified & 5.50 kg tested) and stock stand, respectively (https://www.skogfroverket.no/).

The seeds in Norway were mainly used for the purpose of forestry, timber, and pulpwood. At The Norwegian Forest Seed Centre 79 OECD Forest Seed and Plant Scheme qualified accessions of commercial seed lots of native *Pinus sylvestris* has been stored. *Pinus sylvestris*, as one of the native tree species, has been tested at the Norwegian Landscape Laboratory at NMBU for revegetation and planting purposes in gardens (ornamental) and landscape, so that suitable clones and seed sources can be recommended, and native plant material use in landscaping purposes can be increased. Aroused interest in planting *Pinus sylvestris* in Norwegian forestry necessitates increased seedling production through seed collection coupled with tree improvement in forest stands, or importing genetically enhanced materials, alongside development of models aiming to develop a web tool that would recommend optimal reproductive materials based on specific site conditions (Fjellstad & Skrøppa, 2020). In Norway, the seed production of *Pinus sylvestris* was 100 % from OECD category – source identified amounting 17 kg for the year 2021, while 800,000 plants and 7458 kg seeds of *Pinus sylvestris* were imported in 2021. In Norwegian forestry, 45,502,000 seedlings were delivered in 2021 for forest regeneration, among which around 95.65 % *Picea abies*, 0.5 % others, and 3.85% (1,753,000 seedlings – 28 % unimproved & 72 % genetically improved) *Pinus sylvestris.* Furthermore, the demand for improved materials (seeds and plants) of Scots pine in

Norwegian forestry have increased and mainly fulfilled by import from Sweden, and the improved seeds have been directly sown for plant production in forest, like in Eastern Norway. Additionally, Norwegian Forest Seed Centre has already started a breeding program for Scots pine by developing the Norwegian orchards to produce second generation seed materials in about 15-20 years (Solvin & Sundheim Fløistad, 2023).

Genetic improvement and breeding activities in *Pinus sylvestris* has been initiated in Norway at a lower intensity as compared to *Picea abies* (Norway spruce) and 150 originally selected plus trees were kept as graft in 4 clonal archives for which 7 provenance trials were conducted with a total of 40 provenances, however progeny trials haven't yet been initiated. National tree breeding strategy for 2010-2040 (approved in 2017) guides the breeding programmes in Norway (organized through Norwegian Forest Seed Centre) to produce the bred material with higher survival rates over large areas (climatic adaptation), higher genetic gain in volume production (>20 %), better wood quality and carbon sequestration as well as maintaining genetic variation. Furthermore, there exists a collaborative effort among tree breeding organisations in the Nordic countries, such as the use of a centralized biobank, molecular markers for identifying breeding materials and trait-controlling genes important in selection, adopting "breeding without breeding" approach by integrating genetically evaluated (using molecular markers) trees from regular forests into breeding populations, implementing genomic selection for predicting breeding values, drone technology for tree height measurement in stands and progeny trials, and databases for secure, efficient data management (Fjellstad & Skrøppa, 2020).

The research and development work on forest genetic resources (FGR) are performed mainly by Norwegian Genetic Resource Centre, Department of Forest genetics and biodiversity at NIBIO, Norwegian Forest Seed Centre, along with some national and regional authorities for agriculture and environment, forest sector representatives, research, and educational institutions like NMBU. Likewise, The Forestry Act and the Nature Diversity Act 2009 is the legislation governing forest genetic resources, while the projects are financed either as research grants through the Norwegian Research Council, the Norwegian Agricultural Agency, the Nordic Council of Ministers, and the European Union or by own budget of researching institutions. Also, The Norwegian Forest Seed Centre and NIBIO has maintained a common database of all genetic units tested in FGR research and tree breeding (Fjellstad & Skrøppa, 2020).

1.4 Effect of temperature and precipitation on growth

Scots pine is the most dominant and significantly affected species in European forests by the frequency and intensity of climate change phenomena - extreme weather events like long-term drought, unbalanced wet and dry periods, heat waves and considerable climatic fluctuations. These climatic changes have lots of negative consequences like increased risk of forest fire due to dry period, negative impact on photosynthetic activities, the growth of cambium, resistance against the harmful insects and pathogens, which in turn decreases the diversity and leads to homogenization of forests stands. Also, the dry period affects the radial growth of the

tree, water stress significantly changes the structure of pine stand, and lack of precipitation coupled with high temperature results in die back of pine forests. Likewise, the annual temperature is highly correlated to tree rings growth (width index) rather than the temperature at vegetation season, while annual precipitation is mostly negatively correlated and the precipitation in the vegetation season is highly correlated to tree rings growth (Brichta et al., 2024). For the precise estimation of climate-growth relationship, daily temperature and precipitation data can be used as compared to monthly data, as daily life processes in trees are constantly occurring which could be accounted for in the analysis. The summer drought or frost have instant effect while winter drought has delayed effect on tree rings width, the size and structure of tree rings is affected by many climatic factors among which the drought is the most stressful for trees and narrower rings is caused by droughts. Comparison and analysis of climatic data in extreme years with growth rate of tree found that low temperature (severe) towards end of winter coupled with low sum of precipitation i.e., drought in spring and early summer (vegetation season) limits the growth of tree, inferring precipitation and temperature in March as one of the most influential month in terms of ring width and tree growth (Waszak et al., 2021).

Climatic factors greatly influence the growth of the tree like the radial growth (tree ring width – early wood width and late wood width), and the shortage of water due to higher temperature will reduce the transpiration and photosynthetic activity – tree's capacity to photo-assimilate atmospheric carbon modified, which ultimately will alter tree water potential (decline), secondary growth, xylem structure and functions. Shift from the limitation of low temperature during winter and water shortage due to rise in summer temperature coupled with low precipitation (higher evapotranspiration) have a major impact and control on tree growth. If the availability of water is high plants will uptake more nutrients, a greater number of newer branches and needles will be produced with enhanced photosynthesis and lengthened growing season (Tabakova et al., 2020). Tree height and temperature related variables have the positive correlation, as temperature affects the growth time and rate of tree height – warmer temperature improves the photosynthetic efficiency through enhanced photosynthetic enzyme activities, which increases the material accumulation i.e., more nutrient and carbohydrate accumulation as well as distribution in the trunk. Likewise, warmer temperature cuts the limitations imposed by lower temperature, so the uptake of water and nutrients in root is promoted through enhanced activities of soil microorganisms, the division as well as specialization of cambium and meristems cells is promoted, and these functions encourages the active tree height growth in advance. Furthermore, tree height growth is very sensitive to both current and previous years precipitation and ground water because the prophase of tree height growth is dependent on the increased soil water content and its availability to roots due to increased precipitation from previous autumn to current spring and thawing of soil in spring. However, Scots pine being a pre-season growing species, tree height responses negatively to the precipitation in the previous year's growing season as it limits the photosynthesis due to decreased light intensity and temperature, so that the nutrient accumulation and its availability to top buds growth for current season is limited resulting in lower tree height growth in succeeding year (Y. Zhou et al., 2019). Stomatal density (SD; mm^{-2}) is a key factor that controls the photosynthesis, transpiration, and gas exchange processes in plant leaves, the

number and activities of stomata is found to be also associated with the increased temperature, where higher SD was observed in needle samples from southern end of Europe (warm temperature) as compared to low temperature Northern region (Marek et al., 2022).

Indeed, drought has a huge impact on plant growth, but to objectively quantify its intensity, duration, magnitude, and spatial extent is a difficult task, so standardized precipitation evapotranspiration index (SPEI) is among the widely used drought index calculator that takes the effects of both evapotranspiration and precipitation together into consideration. SPEI is a multi-scalar method for drought assessment based on precipitation and temperature data, which computes monthly climatic water balances (D parameter) - precipitation minus potential evapotranspiration (PET), that has been accumulated at different timescales and helps to identify, analyze, and monitor the droughts events (Pasho & Alla, 2015; Peña-Gallardo et al., 2018). This drought index SPEI calculation was first proposed by (Vicente-Serrano et al., 2010a) in which the monthly climatic water balances simply helps to measure the water surplus or deficit for the month analyzed, using a three-parameter log-logistic distribution those values obtained are transformed to a normal standardized variable, the output is SPEI with average value 0 and standard deviation 1, for direct spatial and temporal comparison. Moreover, SPEI of 0 indicates a value corresponding to 50% of the cumulative probability of D (P-PET), based on a log-logistic distribution and the SPEI values ranges from -5 to +5 in which the smaller the value the stronger the degree of drought and the larger the value the higher the degree of moisture (Y. Ma et al., 2015). Furthermore, the SPEI drought index can be employed into different dry and wet categories as such, extremely wet (EW); SPEI 2.00 and above, very wet (VW); SPEI 1.50 to 1.99, moderately wet (MW); SPEI 1.00 to 1.49, near normal (NN); SPEI −0.99 to 0.99, moderately dry (MD); SPEI −1.00 to −1.49, severely dry (SD); SPEI −1.50 to −1.99, extremely dry (ED); SPEI −2.00 and less (B. Li et al., 2015).

1.5 SNPs markers and SNP-genotyping

Single nucleotide polymorphism (SNP) can be defined as an occurrence of one base variation in a single DNA nucleotide at the specific position in the genome in which the single nucleotide variation results from single-base insertion, deletion, transversion (purine to pyrimidine), and transition (purine to purine or pyrimidine to pyrimidine). Any of the four bases can occur at SNP mutation site, but the SNPs are mainly biallelic with C to T SNPs being most common; these usually form due to spontaneous deamination of 5-methylcytosine to thymine in GC-rich regions. SNPs can be categorized based on their genomic location; cSNPs (gene coding SNPs, relatively rare owing 20% of mutation rate in the surrounding sequence), iSNPs (intergenic SNPs), and pSNPs (perigenic SNPs) and the SNPs being bi-allelic genetic variant (at each SNP maximum 2 alleles and 3 genotypes could be detected), in average occur every hundreds of base pairs (Yang et al., 2020). Further, cSNPs can be categorized into nonsynonymous (a different amino acid is incorporated into the polypeptide by altering the transcribed codon such that, the function of corresponding protein is affected) and synonymous- believed to be biologically silent or inconsequential, but it's impact on gene function (mRNA splicing, mRNA stability, mRNA structure, protein translation and co-

translational protein fold) have also been well documented (Komar, 2009). SNPs act as ideal biological markers because of their higher abundance, wider occurrence, stable heredity, high specificity, extremely low mutation rate and stable genetic variability in the genomes (genes or in regulatory regions), so that genome evolution, genes associated with various traits and their functions can be studied and identified through high-throughput detection and further integration of SNPs easily in genotyping data (J. Zhang et al., 2020).

Various high-throughput SNP genotyping approaches are available which can be compared and utilized based on the requirements (read length, accuracy, throughput), from first-generation sequencing (FGS) based, next-generation sequencing (NGS) based, and third-generation sequencing (TGS) based SNP genotyping to microarray-based SNP genotyping. The main NGS based SNP genotyping platforms to date are based on sequencing-by-synthesis (SBS) and includes, Solexa sequencing (Illumina/Solexa genome analyzer). However, NGS platforms are limited by relatively short read length, so that TGS methods such as Single molecule real time - SMRT (Pacific Biosciences) and Nanopore sequencing (Oxford nanopore technologies) can be considered albeit with reservation, as these platforms also suffer from fidelity issues. Likewise, DNA microarray (DNA chip or Gene chip) based SNP genotyping uses a technique of molecular hybridization of designed probes (two or more) present in a chip with the completely complementary sequences of target gene and in situ fluorescence detection of hybridized products in solid media for SNP analysis, where the base type of targeted sequence is determined based on fluorescence (intensity and type) by laser scanning. DNA microarray can analyze limited initial materials (specific genome regions) and efficiently detect large number of SNPs (known SNPs) with high accuracy in a single experiment (Yang et al., 2020). For instance, in Axiom array, oligonucleotides DNA probes are immobilized on a glass or silicon wafer (array plates) and the amplified genomic DNA is fragmented into 25-125 base pairs (bp) fragments, purified, resuspended, and hybridized to the DNA probes and captured followed by stringent washing for the removal of non-specific background. Now, each captured polymorphic nucleotide on the array surface is queried by the fluorescence labeled solution probes through ligation and differentiated, so that the array can be stained and imaged on GeneTitan MC instrument and data is further handled via software package- The Applied Biosystems™ Axiom™ Analysis Suite (https://www.thermofisher.com), formerly known as the Affymetrix Axiom platform.

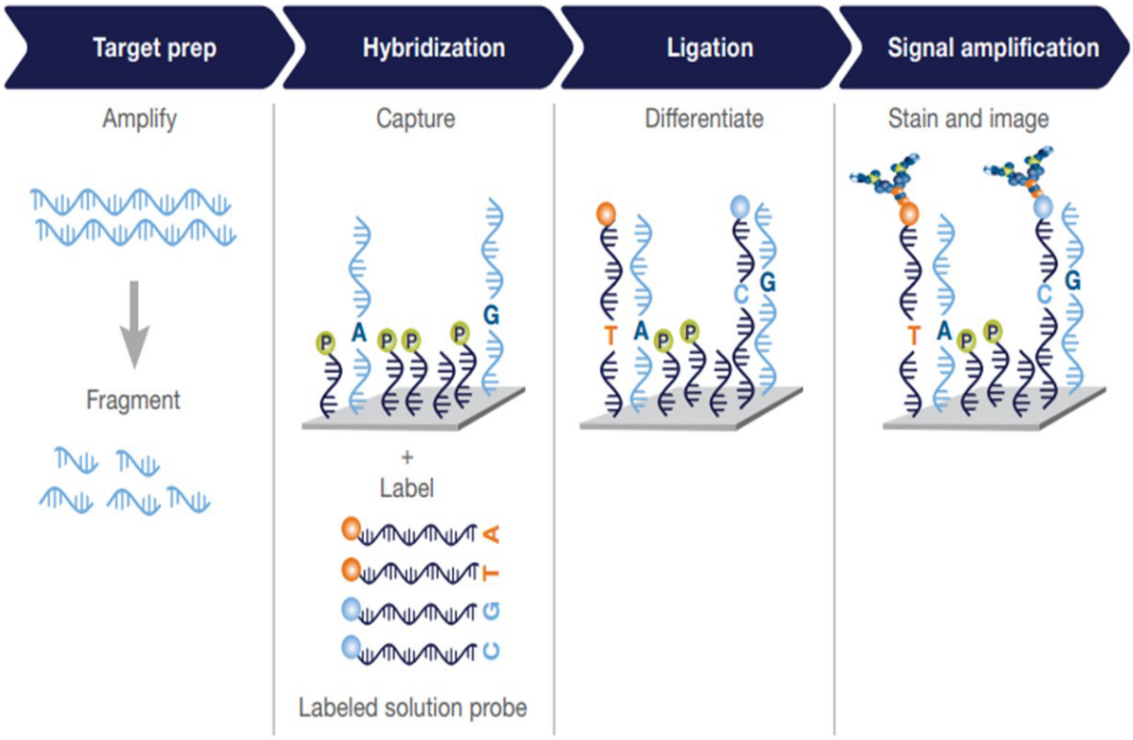


Figure 2. Basic workflow of Applied Biosystems™ Axiom™ microarray (https://www.thermofisher.com).

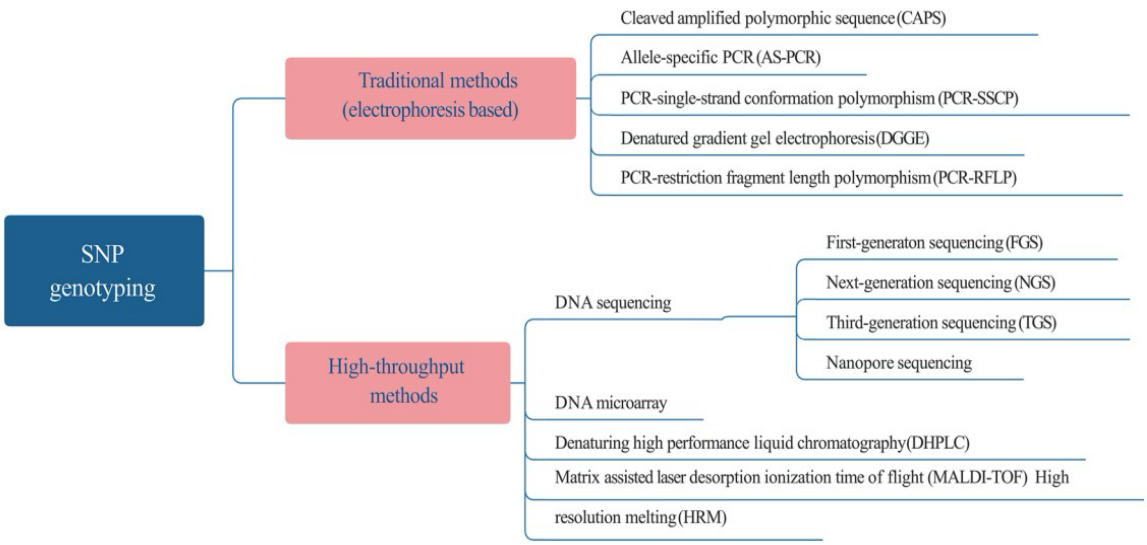


Figure 3. Various SNP genotyping approaches (Yang et al., 2020).

The traditional as well as other SNP genotyping approaches except DNA microarray and DNA sequencing, hardly meets the rapid increased demand for SNP genotyping in large scale due to lower SNPs detection potential and subsequent genotyping despite being time consuming with high device and sample preparation requirements. The DNA microarray or SNP chip microarray is only appropriate for genotyping thousands of SNPs in hundreds of samples as it requires extensive time to develop and design a SNP assay or chip, generates bias to reference genome that hinders to detect unique SNPs, and are also unable to detect and genotype some

SNP loci that are not conserved (occurring in flanking regions of those SNPs) or due to other hits in genome (sequences homology) thus giving false positive or false negative results while SNP genotyping. Moreover, whole genome sequencing (WGS) and resequencing, multiplex PCR and continuous advancement in next-generation sequencing (NGS) technology has altogether made it possible to generate millions of SNPs by screening large part of the genome in different species and some have focused on developing genome wide perfect SNPs harbouring conserved flanking regions and being captured uniquely in the genome during PCR amplification (J. Zhang et al., 2020) . Whole genome resequencing using NGS techniques like restriction-site associated DNA sequencing (RADseq), ddRADseq, and genotyping by sequencing (GBS), through enzymatic digestion, amplification and partial sequencing made it possible to recombine and update partially amplified sequence to the reference genome, so that the unique SNPs can be observed or detected, which was not possible through SNP arrays (Bilginer et al., 2022).

1.6 Genome wise association studies

Grasping the genetic structure of complex traits is essential to have a deep understanding of biology. Majority of traits having agricultural and evolutionary significance are complex traits, controlled by many genetic loci, environment conditions and their interactions (Tibbs Cortes et al., 2021). To establish a strong predictive model for these traits, it is essential to have a significant number of genetic markers analyzed across a large sample size that have precise phenotype, ideally in diverse environments. Nevertheless, advancements in scale, accuracy and affordability of high-throughput sequencing and genotyping are expanding the efficiency of genetic association studies. Specially, availability of economical high-throughput genotyping has become advantageous for the study of non-model organisms, particularly those facing challenges in genome assembly due to genome size and/or complexity (Perry et al., 2022a). Genome-wide association studies (GWAS) is used to identify the associations between genotypes and phenotypes by examining the differences in allele frequency of genetic variants among individuals who share similar ancestry but varying phenotypically (Uffelmann et al., 2021). Genome wide association mapping has the capacity to provide a more accurate understanding of number of genetic regions associated with the trait, ultimately aiding in identification of casual mutations within the mapped regions (Parchman et al., 2012).

In standard GWAS, phenotype and genotype data are gathered from a sizable group of assembled individuals, such as a diversity panel. The genotype data typically includes genome-wide single nucleotide polymorphisms (SNPs), that are identified using methods such as resequencing, genotyping-by-sequencing, or array-based genotyping (Tibbs Cortes et al., 2021). GWAS employs statistical methods for identifying the associations between genomic polymorphisms (such as SNPs, insertions and deletions, and structural variations) and phenotypic variation (Tibbs Cortes et al., 2021; Q. Xiao et al., 2022) and analyze the results to identify a group of genes containing diverse alleles that could be responsible for variation in phenotypes (Burghardt et al., 2017). The effectiveness of GWAS in identifying genes of interest depends on five important factors which includes, genetic complexity of a trait, its

heritability, number of accessions assayed, their relatedness, and the density of the genomic variants. Thus, GWAS is highly potent in identifying a causative gene (i.e., one with segregating alleles contributing phenotypic variation) when there is absence of confounding between relatedness and phenotypic variation, few major loci have significant effects on phenotypic variation, and genetic factors predominantly account for phenotypic variation (i.e., high heritability). Conversely, identifying the genes contributing to phenotypic variation is most challenging when sample exhibit high genetic structuring, traits have low heritability (i.e., phenotypic variation is mostly because of environmental variation), and the genetic basis of the trait is intricate (involving many genes with each having minimal contribution to genetic variance of that trait) (Burghardt et al., 2017).

Moreover, Genome-wide association studies have identified genomic regions linked to various agronomic, physiological, and fitness attributes such as flowering time, plant height, kernel number, stress tolerance, and grain yield. These studies have also explored agriculturally significant traits across a variety of major crop species such as maize (*Zea mays L.*), wheat (*Triticum aestivum L.*), rice (*Oryza sativa L.*), soybean [*Glycine max (L.) Merr.*], sorghum [*Sorghum bicolor (L.) Moench*], barley (*Hordeum vulgare L.*), cotton (*Gossypium hirsutum L.*), and several other crops in addition to the model plant species *Arabidopsis* (Tibbs Cortes et al., 2021). Association studies and genomic selection have been conducted in pine species for various traits including serotiny in *Pinus pinaster* and *Pinus contorta*, circumference, height, and stem straightness in *Pinus pinaster*, oleoresin flow in *Pinus taeda*, and growth as well as wood quality traits in *Pinus sylvestris* (Perry et al., 2022a). Also, the GWAS conducted in *Pinus armandi* have found two potential gene, *CesA2 and CCoAOMT* associated with Diameter at breast height (DBH) (Q. Li et al., 2024). In GWAS of rice, the genes have been found to be associated with geographical differentiation and adaptation during domestication. Additionally, study has identified the genes associated with biochemical and molecular phenotypes such as flavonoid, fatty acid, amino acid, and nucleic acid metabolites (Tibbs Cortes et al., 2021).

Furthermore, GWAS has been used to examine data obtained by high-throughput automated phenotyping. For instance, in Sorghum, GWAS have pinpointed significant associations for panicle architecture through automated feature extraction from images as well as biomass traits through measurements obtained from aerial drones. Genome-wide association studies are used for dual purposes of identifying novel associations with desirable traits and confirming loci discovered by other methods. They can be conducted independently as a part of gene cloning studies or as an initial step in marker-assisted selection among other applications. Indeed, utilizing the information speeds up the progress of crop breeding, for example, genetic loci detected through GWAS on provitamin A levels in maize grain served as the foundation for marker-assisted and genomic selection to enhance this important nutritional trait, thus accelerating crop breeding (Tibbs Cortes et al., 2021).

The advancement in technology and methodology is responsible for the success of GWAS. Various algorithms have been developed for genotype calling from SNP array data with each generation bringing enhancement in accuracy and call rate. New algorithms have also been

developed for calling low frequency and rare variants as well as for inferring haplotypes and structural variants. As GWAS depends more on Whole Genome Sequencing (WGS) data, a corresponding array of tools have been developed to facilitate variant discovery and SNP calling from sequencing data. Similarly, advancements in the statistical imputation of genotypes have been made and the software like PLINK for data management and analysis can effectively process and analyze whole genome SNP array data in a computationally effective method (Tam et al., 2019).

1.6.1 Gene Ontology Analysis

Gene ontology (GO) is a significant step in the field of bioinformatics (T. Zhou et al., 2017), serves as a crucial resource by offering details on a gene molecular function, its involvement in biological process and location within cellular components (Stevenson & Zumajo-Cardona, 2018). Recently, ontologies have emerged as a potent instrument for handling phenotypic data because standardizing terminology across various species and sub-disciplines facilitates inference depending on logical relationship (Oellrich et al., 2015). The GO project has developed three organized sets of terms (i.e., ontologies) describing genes, transcripts and proteins of any organism based on their associated cellular component, biological function and molecular role in a way that is not specific to any species. Collaborative database benefits from utilization of GO terms, which ensures uniform retrievals across them. Additionally, GO terms can be queried at various levels, ensuring annotators to assign characteristics to genes or gene products based on their level of understanding and specificity of the information available about that entity (Botton et al., 2008). An ontology consists of well-defined terms and their relationships, while the structure itself indicates existing representation of biological knowledge and provides a framework for organizing new data. Eventually, an ontology serves as an essential tool which allow researchers to transform data into knowledge (Ashburner et al., 2000).

The clusterProfiler is one of the widely used libraries among Bioconductor packages for gene ontology analysis which offers enrichGO and gseGO functions for over-representation analysis (ORA) and gene set enrichment analysis (GSEA) using Gene Ontology (GO). Further, the AnnotationHub package allows for online querying of GO annotation for non-model organisms by accessing genome-wide annotations from various data providers such as UCSC, Ensembl, NCBI, STRING, and GENCODE. Due to the Bioconductor community's efforts of keeping up to date GO annotation for model and non-model organisms, clusterProfiler facilitates GO analysis across a broader range of species than other tools. Additionally, a GO annotation data frame (for example, obtained data from the BiomaRt or UniProt database using taxonomic ID) can be employed to construct an OrgDb either through the AnnotationForge package or directly via the universal interface for enrichment analysis (Wu et al., 2021).

1.7 DNA methylation, it's significance and applications

DNA methylation is an evolutionarily conserved epigenetic modification that affects gene regulation, gene imprinting, genome structure and integrity which require the addition of a

methyl group to the carbon-5 position of cytosines and contribute to transcriptional repression (Bewick & Schmitz, 2017; Catoni et al., 2018; Gallego-Bartolomé, 2020). DNA methylation in plants is classified into different site classes depending on the sequence context (CG, CHG and CHH), (where H is A,C or T) for which the methylated C (mC) is associated and separate enzymatic pathways are involved in establishment and maintenance of methylation in specific site classes (Bewick & Schmitz, 2017). Three different families of DNA methyltransferases (METases) regulates the process of DNA methylation in plants -which includes maintenance methyltransferases (METs), chromomethylases (CMTs) and de novo domains rearranged DNA methylases (DRMs) (Elhamamsy, 2016). The significance of DNA methylation is not only determined by its abundance, but also by its distribution across different parts of the gene and sequence context such as CG, CHG and CHH. The broad diversity in the type, amount, and location of DNA methylation in a gene during different stages of development has been found in various plant species (Shaikh et al., 2022).

Moreover, the initiation of DNA methylation in plants is regulated by RNA directed DNA methylation (RdDM) pathway, in which the de novo methylation across all sequence context is catalyzed by DNA methyltransferase - domains rearranged methyltransferase 2 (DRM2) (Alakärppä et al., 2018; Gallego-Bartolomé, 2020). In canonical RdDM pathway, the RNA-dependent RNA polymerase 2 is involved in the conversion of single-stranded RNAs (ssRNAs) synthesized from the plant-specific RNA polymerase IV (Pol IV) into double-stranded RNAs (dsRNAs). The synthesized dsRNAs are integrated within Argonaute 4 (AGO4) and Argonaute 6 (AGO6) after slicing into 24-nt small interfering RNAs (siRNAs) by DICER-LIKE 3 (DCL3) enzyme (Gallego-Bartolomé, 2020; He et al., 2011). Involvement of Pol IV to chromatin relies on the histone reader SAWADEE homeodomain homolog 1 (SHH1) and the CLASSY family of putative chromatin remodeler. Another aspect of the pathway involves transcription of non-coding RNAs (ncRNAs) by Pol V, which are aligned through sequence complementarity by siRNA loaded AGO4/AGO6, followed by co-transcriptional slicing. After the formation of AGO-siRNAncRNA-Pol V ribonucleoprotein complex, DRM2 is brought-in to initiate target DNA methylation. Also, involvement of Pol V to the chromatin is facilitated by the DNA methyl-readers SUVH2 and SUVH9, which in turn engage the DDR complex – DEFECTIVE IN MERISTEM SILENCING 3 (DMS3), DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), and RNA-DIRECTED DNA METHYLATION 1 (RDM1) - that is crucial for Pol V's presence on chromatin (Gallego-Bartolomé, 2020). Unlike canonical RdDM, the non-canonical pathways typically participate in establishment of initial DNA methylation at new target loci, such as novel TE insertions, instead of preserving existing heterochromatin (Erdmann & Picard, 2020). The RdDM pathway is guided by small RNAs (sRNAs) originating from various origins such as viruses and Pol II transcripts, and these sRNAs are created from dsRNAs that are cleaved into 21–24nt sRNAs by different DCL proteins. These sRNAs are integrated into different AGO proteins to initiate post-transcriptional gene silencing (PTGS) of complementary RNAs through cleavage and/or translational repression. The specific integration of these sRNAs into AGO4 and/or AGO6 can potentially induce Pol V-and DRM2-dependent methylation of complementary DNA sequences, which could lead to TGS of genes and transcriptionally active TEs (Gallego-Bartolomé, 2020). After the de novo establishment

of methylation via RdDM, maintenance occurs subsequent to DNA replication through methyltransferase 1 (MET1) and chromomethylase 2 and 3 (CMT2 and 3), functioning in the CpG, CHH and CHG contexts, respectively (Gallego-Bartolomé, 2020, 2020; H. Zhang et al., 2018).

In plants, DNA methylation has been identified to be participated in regulating diverse cellular processes such as taking a major part in genome functioning, stability and development, silencing transposable elements (TEs), genomic imprinting and X-chromosome inactivation (Bartels et al., 2018; Shaikh et al., 2022). DNA methylation that occurs in promoter regions is associated with gene expression and eventually involved in regulating growth and development in plants and various other eukaryotes. Methylation of cytosine residues have an impact on binding of DNA with various proteins, including regulatory proteins, which typically prevents the binding of several nuclear proteins which takes participation in transcription and other diverse pathways. Conversely, certain proteins especially bind to methylated DNA sequences and assemble a protein group in the DNA - which is involved in regulation of gene expression. Mostly, cytosine methylation (5mC) in plants is found in transposons and repetitive sequences and inhibits the transcription and transposition, thus is crucial for gene silencing and maintaining genome stability (Shaikh et al., 2022). The methylation of intronic TEs and repeats has also been found to have an influence on mRNA processing mechanisms including alternative splicing and alternative polyadenylation (Gallego-Bartolomé, 2020). In addition to the TEs silencing and repeats, DNA methylation is involved in regulation of various biological processes such as transition to flowering, vernalization, leaf morphology, developmental changes, fertility, floral organ identity, embryonic development, seed development, response to environmental stimuli and genome protection. Recent studies have disclosed that DNA methylation is found to have an essential role in governing important agronomical characteristics, such as heterosis, fruit ripening in tomato and other fleshy fruits, and response to biotic and abiotic stress. Similarly, recent research has also shown the dynamic nature of DNA methylation during embryogenesis and early stages of vegetative growth. The levels of CHH methylation have been found to be higher in embryos as compared to seedlings or adult plants which suggests the crucial role of DNA methylation in embryogenesis. Also, the mutations occurring in *MET1* and *DDM1* have an impact on seed size and development, which shows the significant role of DNA methylation for seed development. Furthermore, experimentally generated hypomethylation in plant genomes has led to several developmental defects, ensuring the essential role of DNA methylation in proper growth and development. Therefore, DNA methylation in plant participates in different processes including fertilization, gametogenesis, vegetative and reproductive development, and interaction between DNA methylation and histone modification (Shaikh et al., 2022).

Indeed, DNA methylation has necessary impact on abiotic and biotic stress response that is critical in plant adaptation mechanisms (Elhamamsy, 2016). Abiotic and biotic stresses in numerous plant species cause variations in the DNA methylation level of particular DNA sequences, leading to changes in the expression of genes related to stress or defense mechanism, ultimately enabling adaptation to environmental stress (Schmidt et al., 2017). The

MET1-mutant *Arabidopsis* plant shows hypersensitivity to salt stress due to significant loss of cytosine methylation and changes in expression of *AtHKT1* gene, which plays a crucial role in plant salt-tolerance mechanism. Exposure of maize roots to cold stress reduces the expression of *MET1* gene and decrease in MET1 level causes demethylation of transposable elements Activator/ Dissociation (Ac/Ds) leading to an activation of abiotic stress-related genes. A similar case of shifting from DNA methylation to demethylation state was also found in rice grown under water stress condition and the demethylation of *Xa21G* gene in rice leads to improved resistance for *Xanthomonas oryzae pv. oryzae* (Elhamamsy, 2016).

Correspondingly, implementation of epigenome editing tools has become advantageous in plant breeding on various levels as it enables the utilization of methylation modifications. Just as the genetic modifications, editing tools could accelerate the domestication process of wild plants by making alterations to the traits that are associated with growth habit, flowering time, yield, nutrition, seed and fruit size, and number in a single generation. However, induced demethylation has a varying effect which can influence both gene expression and repression, as opposed to the gene knockout. Additionally, induced DNA methylation can improve hybrid breeding and plant propagation through tissue culture, which can generate new gene expression patterns in the offspring and aid in managing offspring phenotype. Therefore, epigenome editing could be an effective approach in plant breeding programs for creating new varieties with enhanced agronomic characteristics (Shaikh et al., 2022).

1.7.1 Bisulfite Sequencing

A more cost effective and flexible substitute for Whole genome Bisulfite Sequencing is Bisulfite Sequencing (BS-Seq) in representations of the genome. This method involves using restriction enzyme for genome fragmentation during library preparation and the examples of such method includes RRBS, epiGBS, BsRADseq, epiRADseq, and Creepi (Gawehns et al., 2022). Bisulfite sequencing is a technique used to analyze cytosine methylation which requires treatment of DNA with sodium bisulfite that causes deamination of cytosine to uracil while the methylation at the 5-carbon position prevents this reaction (Gruntman et al., 2008; Henderson et al., 2010). Bisulfite sequencing is followed by various techniques such as sequencing, methylation-specific PCR, combined bisulfite restriction assays, and other methods, so the bisulfite conversion of DNA necessitates prior DNA denaturation, since only methyl cytosines that are situated in single strands are prone to attack. Usually, following denaturation and bisulfite modification, dsDNA is obtained by primer extension, and specific DNA fragment is amplified using PCR (Fraga & Esteller, 2002). The polymerase chain reaction (PCR) of bisulfite treated DNA amplify uracil as thymine while methylated cytosine remains unchanged after amplification. And after amplification, sequencing is performed which allows for determination of the frequency with which sites are present as either cytosine or thymine, serving as an indicator of methyl-cytosine frequency in the original DNA sample. Sodium bisulfite sequencing is a very useful technique when carried out carefully, however, there are chances of some potential pitfalls. The problems are mostly common in plants where methylation can be found in any sequence context. Further, the amplification of unconverted genomic DNA is a frequent problem seen in bisulfite sequencing which is mainly due to incomplete denaturation of the template DNA. This problem can be minimized by designing

primers that are biased towards amplifying fully converted DNA (Henderson et al., 2010). A precise primer design is challenging for successful bisulfite sequencing, which is due to lack of prior knowledge regarding the methylation level and sequence changes after conversion. To maintain unbiased results, it is essential to align cytosine residues at primer binding sites with degenerate bases in primers, while minimizing the number of degenerate positions (Foerster & Scheid, 2010). After the amplification reaction, gel electrophoresis analysis should be performed to verify the attainment of expected size of PCR product, also gel purification is suggested to remove any primer dimers. Afterward, the purified PCR amplified product can be cloned, and the same DNA molecule can be reanalysed using conventional methods (Henderson et al., 2010).

1.7.2 Reduced representative bisulfite sequencing (RRBS) in non-model organisms

The whole genome bisulfite sequencing (WGBS) is regarded as the benchmark for thoroughly analyzing DNA methylation variation by targeting nearly every base in a genome, however, it is limited to species with well-established genomes and is most costly approach for studying genome wide methylation. A more economical option to WGBS is Reduced representative bisulfite sequencing (RRBS), which focuses on a small portion of the genome while still providing a scalable DNA methylation profile with single-nucleotide resolution (Malinowska et al., 2020). RRBS method depends on selecting specific sizes of restriction fragments to create a reduced representation of the genome of a particular strain, tissue or cell type (Meissner et al., 2005). For the species without reference genomes, RRBS can generate accession-specific references solely for the loci that are examined, which can be achieved by incorporating non-bisulfite converted samples in the sequencing (e.g. for bsRADseq) or deducing unconverted reference from the bisulfite treated reads (Paun et al., 2019).

The similarity among individuals in DNA methylation can be measured by integrating the RRBS data, which also provides a ground for describing patterns of epigenetic variation within and among natural populations. The epigenetic measure of such population offers understanding of the distribution of epigenetic variation such as to what extent the natural epigenetic population structure is correlated to underlying genetic population structure and with environmental variation. The studies conducted on natural populations of *Arabidopsis thaliana*, experimental populations of maize, and recombinant inbred lines (RILs) of soybean have reported the higher correlation between epigenetic and genetic structure, suggesting that major epigenetic variation is regulated by genetic control. Moreover, some studies have explained the association of epigenetic population with ecological conditions instead, which suggests direct or indirect contribution of epigenetics to adaptation. This is either because epigenetic changes are caused by environmental conditions and provides a capacity for phenotypic plasticity or continuing heritable responses or/and because the epigenetic variation that are not correlated to overall genetic population structure can be influenced by individual genetic loci. Additionally, high heritability of cytosine methylation in plants provides opportunity to use RRBS loci as markers for linkage mapping purposes in crosses where genetic variation is limited or not present (Paun et al., 2019).

While RRBS could focus on functional areas within a genome, in non-model species, understanding of the genomic context of RRBS sites is usually limited to fragments that share similarities with annotated species. This approach has noted drawbacks because genes and their functions change over time, potentially hindering accurate evaluation of the information obtained in non-model species. Using a reference transcriptome could enhance functional predictions in absence of a reference genome, however, even if a significant portion of RRBS fragments overlap with a well annotated transcriptome or can be annotated with known plant genes, only a few of these fragments overlap with 5' end of gene where increased DNA methylation correlates with gene silencing. The significance of gene body methylation outside of 5' end in plants varies depending on context and taxonomic group, and usually, gene body methylation is weakly or not at all correlated with gene expression. While methylation of the promoter region is strongly associated with silencing, identifying RRBS fragments that overlaps with promoter regions is challenging without a well annotated reference genome of the target species or a close relative. Additionally, it may be difficult to achieve sufficient genome coverage for plants with average to large sized genomes and the identification of differentially methylated regions (DMRs) is another challenge for all RRBS approaches. Studies conducted in several species have reported that methylation changes over large chromosomal stretches (DMRs) are more prone to have an impact on transcriptional activity at nearby loci and contribute towards phenotypic change than at single cytosines (DMPs). Nonetheless, defining DMRs remains challenging even in whole genome analyses. The short fragments interrogated with RRBS techniques (usually< 500 bp) will comprise only a limited cytosine positions and accurately identifying DMRs with statistical confidence from such data will be challenging in many instances. Thus, most RRBS analyses will be constrained to calling DMPs, with their intrinsic randomness (Paun et al., 2019).

Furthermore, the flexibility of RRBS protocols provides various technical ways to enhance effectiveness, for example, selection of restriction enzymes determines which regions of the genome are sampled (Paun et al., 2019). RRBS conducted in maize plant suggest that the RRBS enzymes must be selected carefully by considering variations among genomes and different Region of Interest (ROIs) (Hsu et al., 2017). Methylation sensitive enzymes will focus on sections within plant genomes that show a tendency to avoid densely methylated repetitive regions, resulting in enrichment towards coding regions. However, using these enzymes will increase the likelihood of missing data, especially within DMRs, as methylated individuals at the recognition site will not be included. An enzyme with a restriction site rich in GC content will also bias against repetitive regions because numerous transposable elements (TEs) are associated to regions of genome that are predominantly AT-rich. Additionally, enzymes that cleave more frequently (with shorter recognition sequences) will increase the genome representation irrespective of genomic context. Improvements can also be done by using platform that can generate longer reads (such as MiSeq instead of HiSeq) or by making adjustment to the libraries to obtain information across larger regions using paired-end sequencing techniques (for instance, up to 800 bp for bsRADseq using Illumina). This increase the probability of annotating and detecting DMRs, and potentially identifying promoters (Paun et al., 2019).

1.7.3 Reference free DNA methylation analysis (RefFreeDMA)

There are great possibilities in studying impacts of environment and epigenetic inheritance not only in laboratory but also in natural populations and natural organisms, as animals in the nature usually face the complex evolutionary challenges and ecological adaptation, which cannot be designed in the laboratory settings. Basically, bioinformatics methods are insufficient for analysing sequencing-based DNA methylation data in the unavailability of a high-quality reference genome and in genetically diverse populations in which existing reference genomes could introduce bias into the analysis. An integrated method for analysing DNA methylation at single-base-pair resolution was depicted by combining an optimized high-throughput RRBS protocol with a customized computational method named RefFreeDMA, which enable identification of differential DNA methylation without a reference genome (Klughammer et al., 2015a).

RefFreeDMA is designed as a Linux-based software pipeline, enabling small to moderately sized analyses on a desktop computer (e.g., 40-hr total runtime for completing 20 samples), whereas extensive analyses are effectively parallelized on a computing cluster (Klughammer et al., 2015a). RefFreeDMA generates a deduced genome directly from RRBS sequencing reads, maps the sequencing reads to the deduced genome, performs DNA methylation calling, and detects differentially methylated cytosines and DNA fragments. RefFreeDMA constructs the deduced genome by first grouping the RRBS reads across all samples within a species based on their sequence similarity, which is followed by inference of the consensus sequence for each read cluster. In this consensus sequence, positions containing both cytosines (Cs) and thymines (Ts) among the clustered reads are retained as Cs, since they are likely to represent methylated genomic cytosines and protected from bisulfite sequencing in some but not in all samples. Briefly, all consensus sequences are concatenated with spacer sequences (referred to as stretches of Ns) to make computational tasks easier, which results in construction of deduced genome specifically for a particular species and for the analysis but shared among all samples contributing to the analysis. The subsequent stages of read alignment, DNA methylation calling, and analysing differential methylation follow a similar procedure to DNA methylation analysis with a reference genome. RefFreeDMA addresses significant drawbacks of an existing technique utilizing de novo assembly of MeDIP-seq reads, including low resolution, susceptibility to biases, and lack of quantification. Moreover, it exhibits greater potency and broader applicability as compared to read mapping to the genome of a closely related species, which necessitates a closely matched genome and an unconverted library (Klughammer et al., 2015a).

Additionally, various inherent constraints of reference free DNA methylation analysis were suggested in (Klughammer et al., 2015a) for taking into account while considering this method. Firstly, repetitive elements with high sequence similarity may merge into a single deduced genome fragment, which leads to moderately fewer covered CpGs compared to the reference-based analysis. Secondly, unmethylated cytosines across all samples of a species may not be represented in the deduced genome unless one RRBS sample is sequenced without bisulfite conversion and added to the analysis. Thirdly, the method does not perform de novo assembly of deduced genome fragments, a process that necessitate greater sequencing depth

and broader sequencing coverage than is usually feasible. Consequently, there is possible chances for the same CpG to be included twice in two partially overlapping fragments. The applications of RefFreeDMA includes conducting epigenome wide association studies (EWASs) to explore phenotypic variation within natural populations, examining the epigenetic impacts of various feeds, drugs, and rearing conditions in agricultural research, and meta-epigenome examinations of DNA methylation across entire ecosystems (Klughammer et al., 2015a).

1.8 Aim of the study

The aim of the study was to analyze the SNP genotyping data - GWAS (Genome wide association studies), and the DNA methylation data (analyze output reads from Illumina sequencing of reduced representation bisulfite libraries) using various software, such as PLINK v 2.0, R-stats, and CLC genomic Workbench 22.0.2. The main aim was to determine whether the trees from sowed seeds (Swedish) can be differentiated either genetically or epigenetically from trees derived from wind-blown seeds (Norwegian) from neighboring natural stands of Scots pine. Another aim of the study was to determine whether detected SNP alleles correlate with lifetime growth rate and/or annual growth rates during periods of temperature and/or water stress. Also, the study aimed to characterize the genes associated with the significant SNPs by Gene Ontology analyses. Finally, a secondary aim was to identify whether Swedish vs. Norwegian trees or tall vs. short trees exhibited differential DNA methylation patterns, and if so, whether such differentially methylated regions could be functionally characterized via BLAST algorithms.

2. Material and Methods

2.1 Genotypes and SNPs categorization

Individual sample of *Pinus sylvestris* from 4 different forests – Storvelta (S), Gratvikskogen (G), Brannflata midt (BM), Brannflata Sor (BS) in Norway were used for the study with total of 320 samples, in which 10 sampling surface area from each forest and 8 individual sample trees from each sampling surface were selected. Before sampling, two of these forest, G and S were subjected to harvesting operations by leaving some seed trees. Other two plots, BM and BS, were cleared by fire damage, leaving no remaining seed trees. ALL four of these cleared plots were sown with pine seed purchased from the Swedish Forestry Research Institute by Glommen Mjøsen Skog SA and sold to forest owners who owned BM, BS, G and S. The pine seeds used for sowing were mostly extracted from cones gathered in the Dalarna region of Sweden, but some seeds were taken from cones gathered on the Norwegian side of the border, directly east of Gratvikskogen and Storvelta. One of the objectives of this study was to genetically differentiate between the Swedish origin trees (sown seed) and naturally occurring population (Norwegian).The phenotypic data in terms of total height, diameter, age of trees from 5 to 10 years, height increment in each year i.e., from Year 2011 to 2020 was recorded, and the origin of trees either Swedish or Norwegian was inferred from the SNP data and STRUCTURE analysis. Ten to thirty needles were harvested from each tree, placed in an Eppendorf tube, and immediately flash frozen and stored in liquid N_2 before transferring to a -80 °C freezer. The needles were then freeze dried. The genomic DNA from all the samples were extracted by BioBank AS (Hamar, Norway) and SNP genotyped using the Affymetrix platform and the PiSy50K array developed at the University of Helsinki (Kastally et al, 2021). From all the samples, data for 47712 SNPs were obtained, and 10 individual samples failed the quality control and 10 were missing data, so the total of 300 out of 320 samples were used for the downstream analyses. The final SNP call file (.txt) from Affymetrix containing information for those 47712 SNPs were assigned to either of six classes which includes, Poly High Resolution, Mono High Resolution, No Minor Homozygotes, Call Rate Below Threshold, off target variants (OTV), and Other. Finally, only 42920 SNPs falling in category Poly High Resolution and No Minor Homozygotes (NMH) were used for further analysis and were converted to pgen, pvar, and psam file using the command in Plink 2.0 with the help and suggestions of Chedly Kastally.

2.2 Population structure and admixture analysis

Using STRUCTURE software v.2.3.4, the population structure analysis was performed through the series of steps by creating a separate project. Initially, the input file was created with the names of SNPs marker in the row and individual ID included in the column. After clicking the option – file, the project directory and name of the project was then specified and the selection of input file for the analysis was completed. Likewise, various input file characteristics, such as the number of individuals (310), number of SNPs loci (42920), and

missing data value (-9) was also specified in STRUCTURE. v.2.3.4. The analysis was performed selecting the admixture model with two replicates of K values ranging from 1 to 10, burn-in period of 10000 and 20000 Markov chain Monte Carlo (MCMC) replication after burn in. Thereafter, to parallelize the runs of STRUCTURE, Structure threader program was used by creating a virtual environment in Python. Further, the output from STRUCTURE v.2.3.4 was used as an input and processed in STRUCTURE HARVESTER executing "Evanno method" to determine the best K value through statistics Δ K (Earl & vonHoldt, 2012). For the better visualization of true number of populations, STRUCTURE v2.3.4 and fastSTRUCTURE were used to generate the admixture plot and presented in the result section.

2.3 Genome Wide Association Study (GWAS)

Genome wide Association Study was conducted in Plink v 2.0 to associate SNPs with phenotypic data. Based on the phenotype, two different association study was performed for 300 samples of *Pinus sylvestris*– (i) Association study of SNP genotype associated with tree height (Tall or Short) based on Origin of tree i.e., Swedish (S), Norwegian (N), and Total - Swedish & Norwegian (Total - N+S) and (ii) Association study of SNP genotype associated with tree growth (good or bad) based on temperature and precipitation. For the association study of tree height, all the samples were separated in terms of their age and tall versus (vs) short trees were sorted based on total height of individual trees for their respective ages - by calculating average value and standard deviation (std). The sorting process was done separately for the origin of tree – N and S, and for total (N+S), then the height of the trees that falls 1 Standard deviation above the Mean Value was categorized as Tall whereas the height falls 1 Standard deviation below the Mean Value was categorized as short trees. Subsequently, the three different psam files were created for each of the Swedish, Norwegian, and Total (N+S) trees including Individual ID, Family ID and tree height denoted as 1 for tall, 2 for Short, and -9 for trees belonging to neither of them. Also, the covariance file (covar.txt) was created considering the four different sampling locations – Storvelta, Gratvikskogen, Brannflata midt (BM), Brannflata Sor (BS) as covariance and integrated in the analysis. After that the association analysis was performed for N, S, and total (N+S) in the PLINK 2.0 software using the flag –glm and –assoc. Association based on dominant and recessive allele was performed-for with and without covariance, as well as for interaction, while based on Genotypic, the association was performed only for without covariance and interaction. Thus, the SNPs significantly (P-value less than or equal to 0.01) associated with tree height (Tall or Short) based on origin of tree i.e., Swedish (S), Norwegian (N), and Total - N+S were identified.

Moreover, for the association of growth with seasonal temperature and precipitation, the Standardized Precipitation Evapotranspiration Index (SPEI) was calculated using R-studio to detect the period with a water deficit. The homogenized monthly mean temperature and precipitation data from 2011 to 2020 was taken from Flisa II station and potential evapotranspiration was estimated using the Thornthwaite method as described by (Vicente-Serrano et al., 2010b) and the SPEI was calculated (figure 4) on 3 month scale (SPEI-3). To

analyse the association of growth, two different time-period of the years were selected which includes, dry summer months from June to August and dry spring months from March to May. All the 300 individual sample of trees were separated based on their age and the growth was determined in each year from 2011 to 2020 using the height increment for each of those year. The similar method of calculating the average and standard deviation was used as applied in association study of height based on origin of trees. Likewise, the growth of trees either good or bad was identified for each of the year considering 1 Standard deviation above the mean value and 1 Standard deviation below the mean value representing good growth and bad growth, respectively. Depending on the calculated SPEI value-integrated over three-month scale, only the year which showed the high and low SPEI value was chosen for each time-period. For Dry summer (June-August), the High SPEI and Low SPEI value was observed in the year 2018 and 2012, respectively. While, for the period of dry spring (March-May) High SPEI and Low SPEI value was observed in the year 2017 and 2019, respectively. Also, the High SPEI was named as Low Temperature High Precipitation (LTHP) and Low SPEI as High Temperature Low Precipitation (HTLP).

Finally, the four different psam files were created for these four different categories identified. The phenotype for individual tree in the psam file was denoted as 1 for good growth, 2 for bad growth, and -9 for the samples that were neither of them. After this, the association analysis was performed in PLINK 2.0 (--glm and --assoc) only for without covariance in the dominant, recessive, and genotypic categories, thus the significantly associated SNPs were identified by sorting the P-value less than or equal to 0.01. Using Venn Diagram, the significantly associated SNPs that were identified in both association study were shown for better visualization.

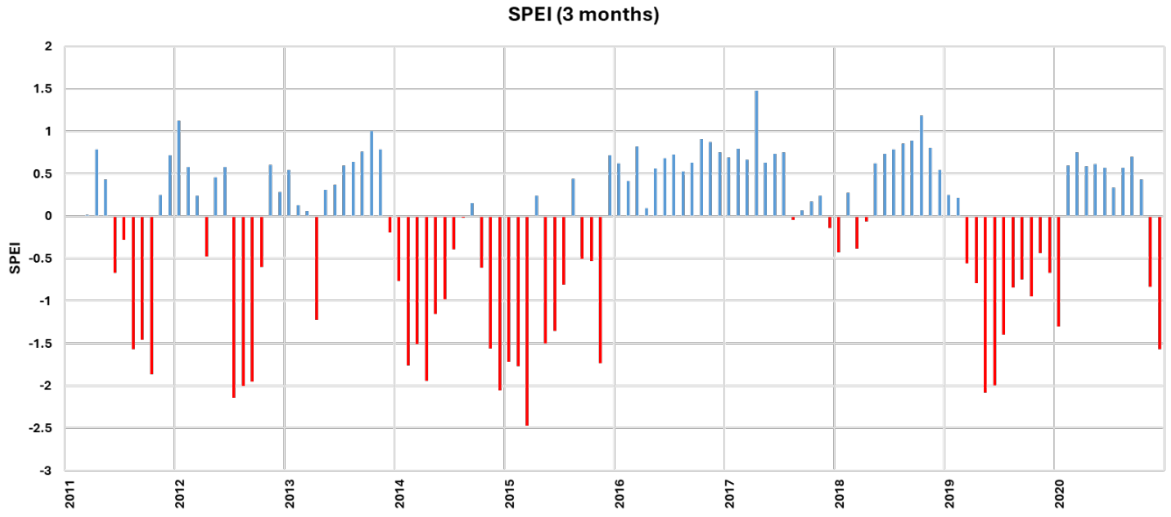


Figure 4. SPEI calculation over the 3 months scale (SPEI 3) based on the homogenized monthly mean temperature and precipitation data from 2011 to 2020 taken from Flisa II station.

2.4 Gene Ontology Analysis

Gene Ontology (GO) Analysis was performed using R-studio to identify the Biological Process (BP), Molecular Function (MF), and Cellular Components (CC) involved. Using the

Metadata file (47712 SNPs blast searched with nucleotide sequences of *Pinus sylvestris* as well as with transcriptome sequences of *Pinus taeda*), the SNPs that were found to be significantly associated with the phenotype in the association study were compared - to identify the genes associated with those significant SNPs, but not all the significant SNPs were associated with the genes. The genes that were identified in both association analysis based on dominant, recessive, and genotypic were utilized to conduct Gene Ontology analysis. The list of gene symbols was used as input file to carry out the analysis in R-studio using enrichGO function provided by ClusterProfiler package from the Bioconductor and the AnnotationHub package of Arabidopsis thaliana was used. Consequently, the redundant terms from the list of enriched GO terms were eliminated using the function in the ClusterProfiler.

2.5 Differential DNA methylation Analysis

Purified genomic DNA from 15 Swedish and 15 Norwegian trees, roughly 50:50 tall vs. short was shipped to CD Genomics in the USA for generating reduced representation bisulfite sequencing (RRBS) data. Briefly, the DNA samples were digested with restriction enzymes – DpnII & MspI for RRBS library preparation, in which the short fragments containing methylated dinucleotides at the ends are generated, end repaired, and size selected (300 bp) and adapter-ligated. The DNA fragments are then subjected to bisulfite conversion (sodium bisulfite), PCR amplified, and sequenced by Illumina paired end (150 bp - each end) sequencing platform. Following QC and trimming, sequencing reads were pre-processed using RefFreeDMA (Klughammer et al., 2015b) basic guidelines by splitting the scripts in both python and R-studio. The two binary variables, Norwegian vs. Swedish and Tall vs. Short were used for conducting DNA methylation analysis in three different sequence contexts - CpG, CHG, and CHH. While performing the analysis, the sequence reads were processed separately. The differentially methylated fragments and differentially methylated motifs for Read 1 and 2 of both binary variables (Norwegian vs. Swedish, NvS, and Tall vs. Short, TvS) in three different sequence context- CpG, CHG, and CHH (3 context * 2 variables * fragments & motifs = 12 files, each for Read 1 and Read 2) were generated. And filtered in steps - P-adjusted value less than or equal to 0.05 ($P \leq 0.05$), followed by coverage greater than 8 and less than 200 filtering value in meth cov g1 (i.e. comparative sample 1) and in meth cov g2 (i.e. comparative sample 2) - to identify the significantly methylated sequences. Further, filtering was also done in the meth.mean.diff and diff.mean.meth column for differentially methylated fragments and differentially methylated motifs, respectively with same value range for both Read 1 and 2 (in same variable and context). For the comparisons of NvS and TvS, the meth.meth.diff values for fragments and diff.mean.meth values for motifs were standardised for all contexts (CpG, CHG, CHH) and same values were applied for both reads (1 and 2). The standardized parameters values applied in NvS for fragments were - CpG (meth.mean.diff of +/- 15), CHG (meth.mean.diff of +/- 30), and CHH (meth.mean.diff of +/- 7) and for motifs were - CpG (diff.mean.meth of +/- 50), CHG (diff.mean.meth of +/- 50), and CHH (diff.mean.meth of +/- 50). Furthermore, the standardized parameters values applied in TvS for fragments were - CpG (meth.mean.diff of +/- 20), CHG (meth.mean.diff of +/-

30), and CHH (meth.mean.diff of +/- 7) and for motifs were - CpG (diff.mean.meth of +/- 20), CHG (diff.mean.meth of +/- 20), and CHH (diff.mean.meth of +/- 50).

After the filtering step, MultiBlastx and Multiblastn was performed in CLC Genomics Workbench 22.0.2 and the parameters for blast search was set (Appendix). MultiBlastn against the created blast database of *Pinus jeffreyi* (Accession number: SAMN12121614) and MultiBlastx against the Refseq protein database was conducted for both the differentially methylated fragments and differentially methylated motifs. MultiBlastn of binary variable (NvS) for different context was performed separately, with read1 and read 2 together for both motifs and fragments. Similarly, MultiBlastn for TvS differentially methylated motifs was performed for read 1 and read 2 together, but for fragments it was done separately for read 1 and read 2. Furthermore, MultiBlastx of binary variable for different context was performed separately for read1 and read 2 - motifs (NvS as well as TvS) and fragments (TvS), while for NvS differentially methylated fragments it was performed together for read 1 and read 2. Finally, the differentially methylated sequence (motifs and fragments) and the results from both MultiBlastn and MultiBlastx were analysed.

3. Result

3.1 Genotypes and SNPs categorization

Genotypes for a total of 47712 SNP markers were obtained for 310 *Pinus sylvestris* DNA samples, after quality control, using the PiSy50K SNP array based on the Affymetrix Axiom platform (Kastally et al., 2022). The SNPs markers were categorized into different classes, of which only 42920 SNPs from the category Poly high resolution (PHR) and No-minor homozygote (NMH) were used for further analysis. Out of total number of SNPs (42920), the highest percentage (97.9%) were in HWE (P-value ≥ 0.01) (figure 5A). Moreover, the number of SNPs above MAF ≥ 0.05 was 41974 (figure 5 B) and the number of SNPs falling into different MAF values is shown in figure 6. The observed heterozygosity, O (HET) was higher in majority of SNPs as compared to expected heterozygosity, E (HET) (figure 7).

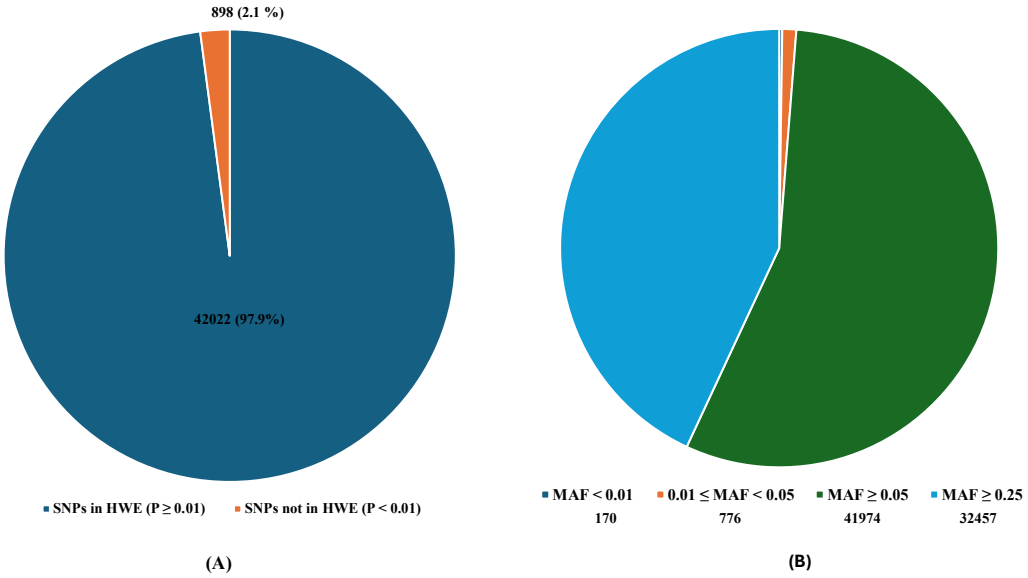


Figure 5. Categorization of SNPs either in HWE (P ≥ 0.01) or not in HWE (P < 0.01). (B) Categorization of SNPs in different MAF categories.

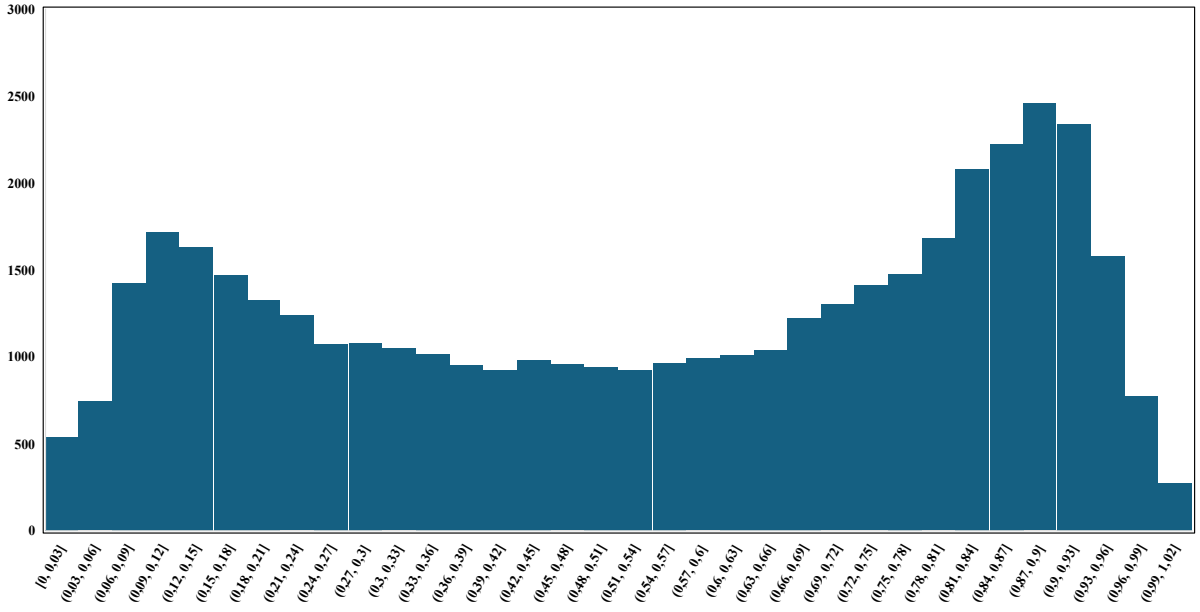


Figure 6. Number of SNPS falling into different MAF categories.

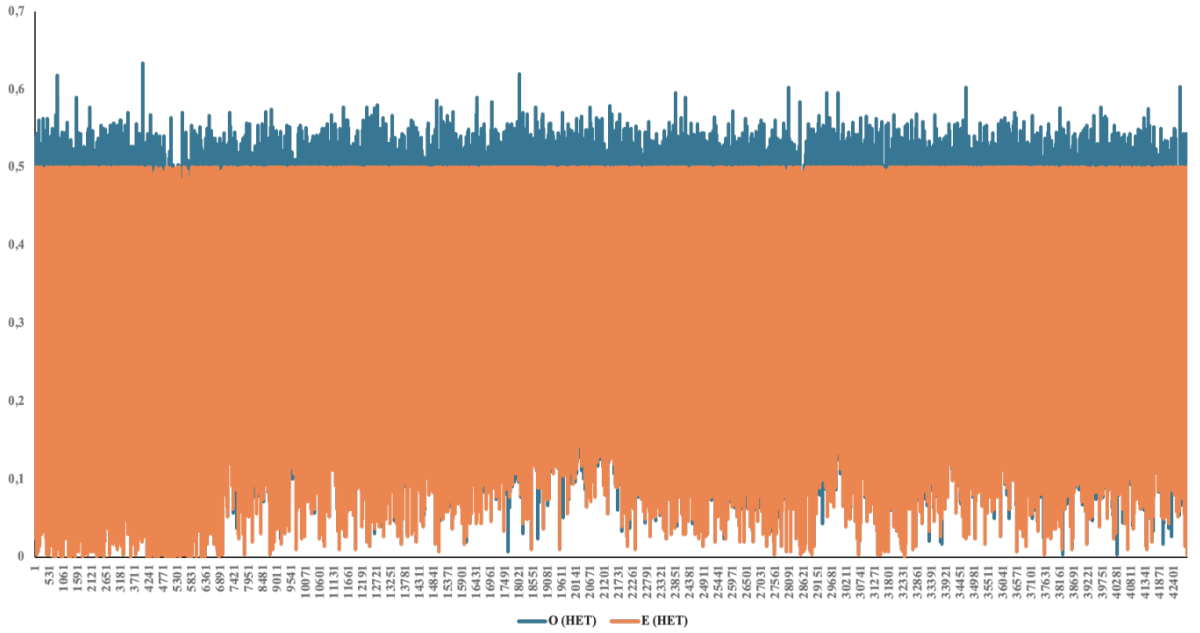


Figure 7. Observed and expected heterozygosity for each 42920 SNPs obtained.

3.2 Population structure and admixture analysis

Among four forest plots, Branflatta Midt (BM), Brannflatta Sør (BS), Gratvikskogen (G) and Storvelta (S), two of these forest, G and S were subjected to harvesting operations by leaving some seed trees before sampling. Other two plots, BM and BS, were cleared by fire damage, leaving no remaining seed trees but there still have been some viable Norwegian seeds left in the soil that germinated along with the sown Swedish seeds. Among the 8 different population,

one large population was identifying as the Swedish population (population 4), while other represented different mixtures of Norwegian and Swedish. Particularly, Gratvikskogen showed the fewest individual (population 4 - Swedish) and belonged largely to population that were more mixed with the Swedish (Table 1). This was likely due to the proximity of the cones (containing seeds) collection sites on the Norwegian side closer to Gratvikskogen than the other plots.

Analysis of 42920 SNPs in STRUCTURE V.2.3.4 was mainly performed to determine the true K - value (the true number of populations) using an admixture model and to analyse various characteristics of populations such as allele frequency divergence, mean FST and expected heterozygosity between individuals. STRUCTURE HARVESTER was used to process the output results of STRUCTURE program executing "Evanno method" as a mean likelihood value - LnP(k) for each K-value and plots were generated. A plot of mean likelihood as a function of K visually indicated that the likelihood increased until the true K was reached and then plateaued (Figure 8). The ΔK plotted for each k value showed a modal value for the distribution and generates the true k value which was eight clusters (Figure 8). The admixture plot was generated from the fastSTRUCTURE and STRUCTURE V.2.3.4 which showed the population structure graph at different k-value and at true K- value i.e., eight distinct population (8 clusters) in *Pinus sylvestris* (Figure 9 and Appendix figure A2 & A3).

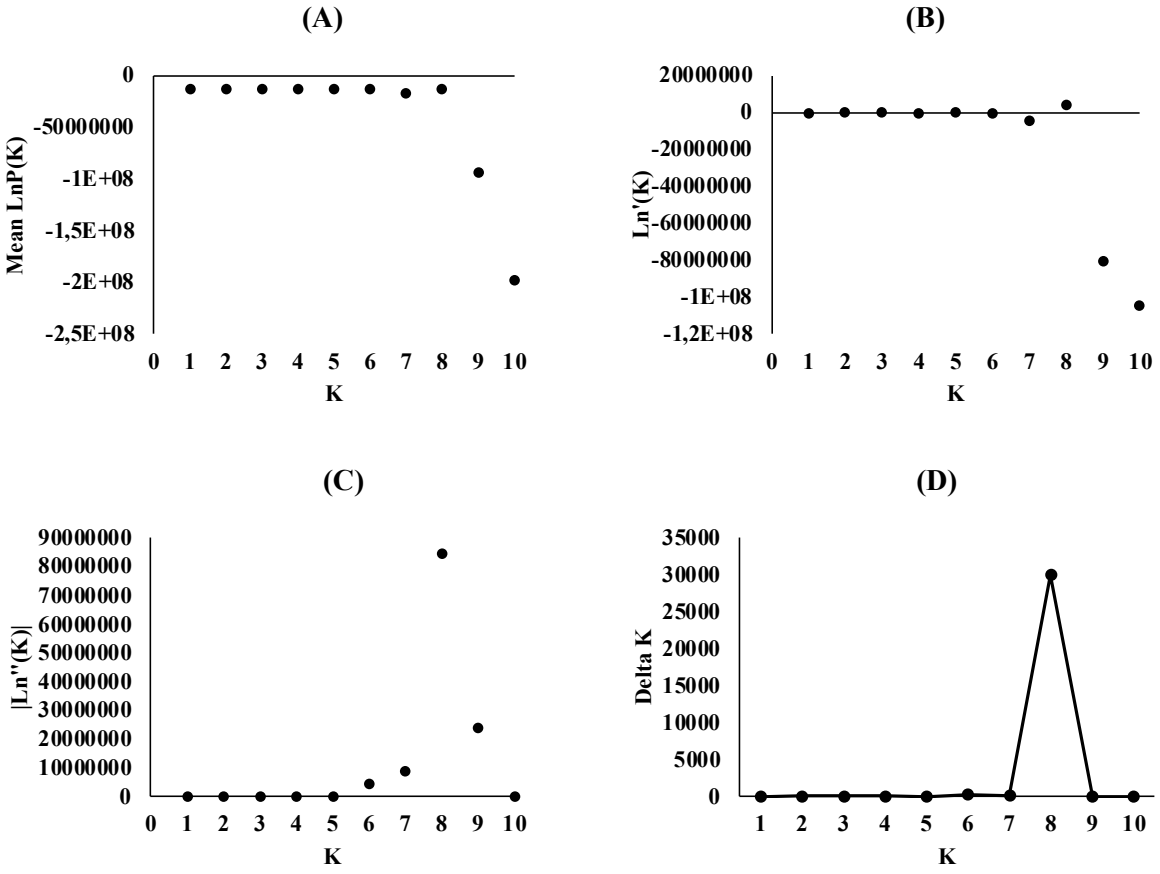


Figure 8. shows a graphical representation of the steps used to identify the true number of populations (K) from STRUCTURE analysis on a dataset of 310 Scots Pine individuals genotyped at 42,920 SNP loci. *Panel A - Mean likelihood plot LnP(K) for each K value over 10 runs. Panel B - First order rate of change of LnP(K) with respect to K value, Ln'(K) = LnP(K) – LnP(K-1). Panel C - Second order rate of change (absolute value) of LnP(K) as per K value, |Ln''(K)|=|Ln'(K+1)-Ln'(K)|. Panel D - Delta K (ΔK) value, calculated using the formula ΔK = mean (|Ln''(K)| / sd [LnP(K)]), plotted against K values to provide a modal value indicating the true K or the highest level of structure. In this study, the true K is determined to be eight clusters.*

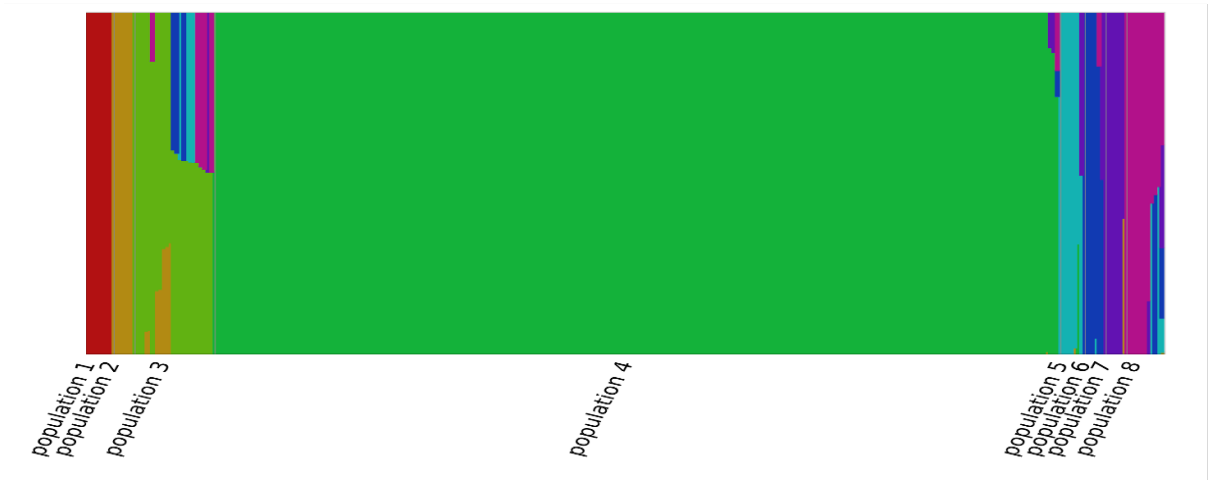


Figure 9. An admixture plot from fastSTRUCTURE shows eight populations, represented by different colors, at the optimal K value (K=8).

Table 1. Estimation of sampled trees from each of the forests belonging to different populations. N: Norwegian, S: Swedish

	BM	BS	S	G	Total
Pop 1 (N)	3		4		7
Pop 2 (N)		2	1	3	6
Pop 3 (N)				20	20
Pop 4 (S)	69	74	75	22	240
Pop5 (N)				6	6
Pop 6 (N)				5	5
Pop 7 (N)				6	6
Pop 8 (N)				10	10
Total	72	76	80	72	300

The proportion of membership of sample observed in each cluster of *Pinus sylvestris* ranged from 0.021 (cluster 7) to 0.759 (cluster 4). The genetic diversity between individual within same cluster showed that the average distance varied from 0.1836 (cluster 8) to 0.3191 (cluster 4). The mean Fst value reported in all the cluster ranged from 0.0118 (cluster 4) to 0.5253 (cluster 8) which showed the remarkable genetic differentiation in all the cluster, except cluster 4 with low genetic distinctness within cluster (Table 2).

Table 2. Estimation of cluster proportions, expected heterozygosity, and mean Fst values using STRUCTURE in Scots Pine.

Inferred Clusters	Proportion of membership of sample in each cluster	Average distances (expected heterozygosity) between individuals in same cluster	Mean value of Fst
1	0.038	0.2198	0.4771
2	0.025	0.2161	0.4716

3	0.06	0.2593	0.2594
4	0.759	0.3191	0.0118
5	0.025	0.2081	0.4714
6	0.027	0.2482	0.3245
7	0.021	0.1978	0.5142
8	0.046	0.1836	0.5253

The allele frequency divergence among population showed that the highest nucleotide distance was observed between cluster 4 and 6, while the cluster 5 and 8 were found to be closely related with lowest (0.0299) net nucleotide distance between them (Table 3).

Table 3. Allele-frequency divergence (Net nucleotide distance) among Scots Pine populations, computed using point estimates of P.

Inferred Clusters	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	-	0.0922	0.1206	0.1253	0.0614	0.1294	0.1144	0.0883
2		-	0.0913	0.0948	0.0319	0.1019	0.0861	0.0599
3			-	0.1211	0.061	0.1303	0.1161	0.0882
4				-	0.0648	0.1323	0.1229	0.0935
5					-	0.071	0.0569	0.0299
6						-	0.127	0.0985
7							-	0.0855
8								-

3.3 Genome wide association study (GWAS)

The study involved two separate investigations aimed at understanding the genetic factors influencing Plant Height (PH) and Plant growth through Genome Wide Association Studies (GWAS). In the first study, our goal was to identify SNPs significantly associated to PH across population of Scots pine. The second study aimed to examine the potential relationships between SNPs markers and annual Plant growth, considering variations in temperature and precipitation across different seasons.

3.3.1 Identification of SNPs associated with Plant Height based on their Origin

Genome Wide Association Study (GWAS) was conducted to identify the significantly associated SNPs with plant height (PH) in Norwegian, Swedish and combined population (Total(N+S)) of Scots pine from 4 forest plots- Brannflata Midt (BM-), Brannflatta Sør (BS), Gratvikskogen (G), and Storvelta (S).In the Norwegian sample, significantly associated SNPs with PH were identified in dominant alleles (53 SNPs) and recessive alleles (1 SNP) by considering the covariance. Without covariance, the significant association included 84 SNPs with dominant alleles, 19 SNPs with recessive alleles, and 49 SNPs with genotypes (Table 4). For Norwegian trees, inclusion of interaction association model did not exhibit any SNPs marker associated with PH in any forest. In terms of Swedish population, significantly associated SNPs with dominant alleles were 217 and recessive alleles were 32 when considering covariance. Without covariance, the associations expanded to include 279 SNPs associated with dominant alleles, 94 SNPs with recessive alleles, and 265 SNPs with genotype. Interaction analysis identified 98 SNPs (49 in BS and 49 in S) associated with dominant alleles and 2 SNPs (in BS) associated with recessive alleles, whereas no SNPs were associated with genotype (Table 4). The predominance of dominant allele was observed in both, Norwegian and Swedish trees, and dominance of dominant allele was much higher when the covariance was not included.

For the combined population of Norwegian and Swedish (Total (N+S)), significant associations were detected with 344 SNPs identified in dominant allele and 56 SNPs identified in recessive alleles when covariance was considered. Association analysis without covariance included 313 SNPs identified in dominant alleles, 131 SNPs in recessive alleles, and 299 SNPs in genotype. Based on interaction, association analysis revealed 585 SNPs (150 in BS, 244 in S, and 191 in G) associated with dominant alleles, 30 SNPs (14 in BS, 5in S, and 11 in G) with recessive alleles, and 106 SNPs 8 26 in BS, 52 in S, and 28 in G) with genotype interaction (Table 4). In BS, interaction association model identified 150 significant SNPs associated with PH by dominant model, and 14 significant SNPs were associated with PH by recessive. Additionally, 10 significant SNPs were associated with PH by an additive genetic model, and 16 significant SNPs were associated with PH by dominant deviation genetic model. Likewise, in forest S, analysis revealed a substantial genetic influence on PH with dominant model identifying 244 significant SNPs. Furthermore, recessive genetic model identified 5 significant SNPs associated with PH, while additive genetic model (20 SNPs) and dominant deviation model (32 SNPs) also showed influence on PH. In G, a dominant model exhibited 191

significant SNPs associated with PH, recessive model revealed 11 significant SNPs, while additive genetic model and dominant deviation model identified 7 SNPs and 21 SNPs, respectively.

Table 4. Significant SNPs identified based on Origin of 300 Scots Pine (Pinus syvestris) samples.

Origin	Dominant	Dominant Interaction	Dominant without covariance	Recessive	Recessive interaction	Recessive without covariance	Genotypic interaction	Genotypic without covariance
Norwegian								
Significantly* associated SNPs	53	0	84	1	0	19	0	49
Swedish								
Significantly* associated SNPs	217	98	279	32	2	94	0	265
Total (Norwegian + Swedish)								
Significantly* associated SNPs	344	585	313	56	30	131	106	299

* Significance determined at p-value ≤ 0.01

Table 5. Significantly associated SNPs with plant height (PH) identified in different forest plots.

Forest plots	Dominant model * covariant interaction	Recessive model * covariant interaction	Genotype model * covariant interaction
BM	-	-	-
BS	199	16	26
S	293	5	52
G	191	11	28

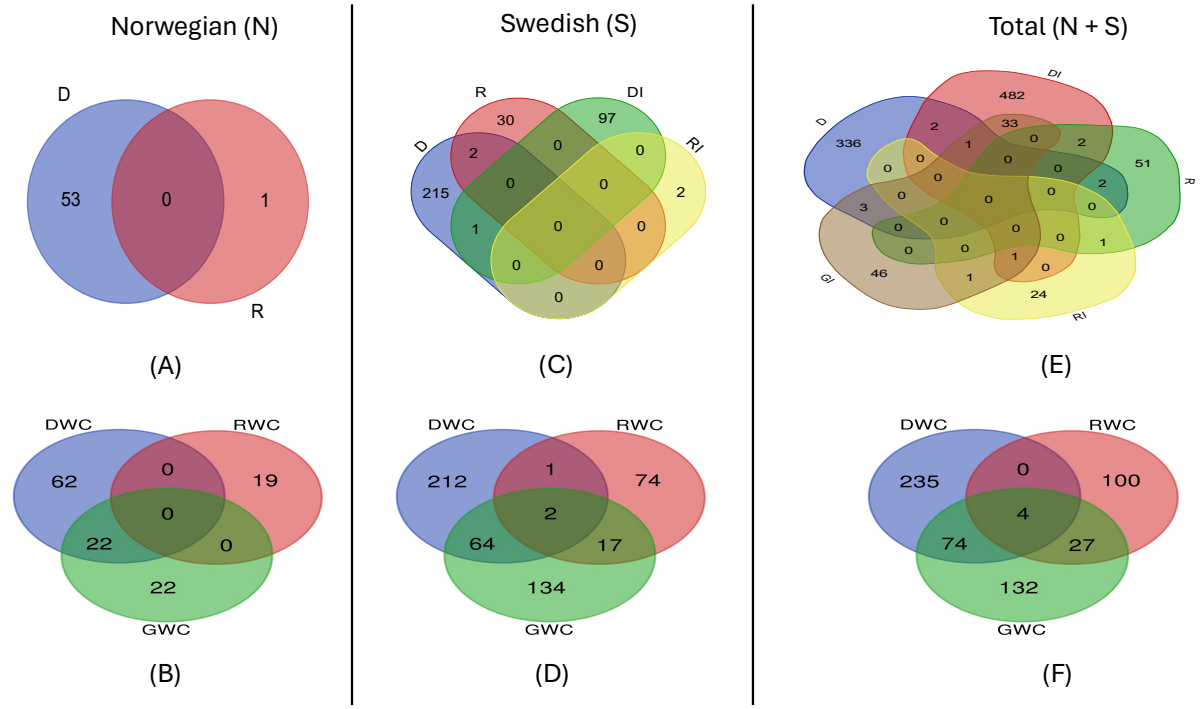


Figure 10. Venn diagram showing significantly associated SNPs based on the origin of 300 Scots Pine samples*. (A) represents significantly associated SNPs based on the dominant - D, recessive - R and dominant interaction - DI, of Norwegian trees. (B) represents significantly associated SNPs based on the dominant, recessive and genotype without covariance (DWC, RWC & GWC, respectively) of Norwegian trees. (C) represents significantly associated SNPs based on the dominant, recessive, dominant interaction, and recessive interaction- RI of Swedish trees. (D) represents significantly associated SNPs based on the dominant, recessive and genotype without covariance of Swedish trees. (E) represents significantly associated SNPs based on the dominant, recessive, dominant interaction, recessive interaction, and genotypic interaction - GI of Total(N+S) trees. (F) represents significantly associated SNPs based on the dominant, recessive, and genotype without covariance of Total(N+S) trees.*

3.3.2 Association study of SNPs with Annual Plant growth under varying temperature and precipitation

Understanding the genetic basis of annual plant growth and its response to changing environmental conditions is important for elucidating adaptive strategies of plant. GWAS helps to understand the genetic architecture of complex traits like plant growth.

In our investigation of the genetic factors influencing yearly Plant growth, we conducted a detailed re-evaluation of our dataset to identify SNPs markers that specifically correlate with Plant growth during different time period with seasonal changes in temperature and precipitation. Based on the Standardized Precipitation Evapotranspiration Index (SPEI), which provides a standardized measure of drought severity by integrating both temperature and precipitation data. Four different time periods were selected to identify SNP markers associated with growth during seasons varying in temperature and precipitation. During the period of March-May, the highest number (376) of significantly associated SNPs was identified in the dominant allele model, while lowest number (88) was identified for the

recessive allele model, in the Low SPEI category (HTLP, 2019). Conversely, the High SPEI (LTHP, 2017) marked the highest number (982) and lowest number (400) of significantly associated markers based on the recessive allele and dominant allele models, respectively. Based on the genotypic model the number of significantly associated SNPs were 292 and 351 for Low SPEI and High SPEI, respectively (Table 6).

Furthermore, both the SPEI category of the period June-August (year 2012 & 2018) showed the greater number of significant SNPs for dominant allele model while the lowest number was found for recessive allele model. The number of SNPs identified in the Low SPEI were 55 for dominant allele, 34 for the recessive allele, and 47 for the genotypic. For the high SPEI category, 488, 111 and 387 significant SNPs were identified by the dominant, recessive, and genotypic association, respectively (Table 6).

Table 6. Significant SNPs associated with growth and with seasonal temperature and precipitation for different time-period.

Time period with SPEI	Dominant without covariance	Recessive without covariance	Genotypic without covariance
		March-May	
HTLP (Low SPEI), 2019			
Significantly* associated SNPs	376	88	292
LTHP (High SPEI), 2017			
Significantly* associated SNPs	400	982	351
		June-August	
HTLP (Low SPEI), 2012			
Significantly* associated SNPs	55	34	47
LTHP (High SPEI), 2018			

Significantly* associated SNPs	488	111	387

* Significance determined at p-value ≤ 0.01

3.4 Gene Ontology Analysis

3.4.1 Gene Ontology (GO) analysis of genes associated with Plant height

Understanding the biological importance of genes identified through GWAS is crucial for elucidating the underlying mechanisms regulating these traits. Gene Ontology (GO) analysis provides a systematic approach for annotating genes based on their biological processes (BP), molecular functions (MF) and cellular components (CC), thus facilitates interpretation of GWAS results. Notably, only a subset of the SNP marker on the array were annotated as lying in " known " sequences identified via Blast searches against various plant sequences databases. The subsequent GO analysis was performed exclusively on this subset of associated SNP marker for which annotation information was available.

Among the 29 genes identified in Norwegian and 172 in Swedish trees, only 3 genes were found to be common between them (Appendix A5). The *CUL1* and *RPK2* genes were identified from the SNPs (7) that were found to be commonly associated with Norwegian and Swedish individuals in GWAS. However, the *CPL1* gene was identified as common to both populations, although identified from different SNPs. Considering the GO analysis, the *CUL1* gene identified from the Swedish trees was found to be associated with the recessive allele model or identified from the SNPs that were significantly associated based on the recessive allele. While in the Norwegian sample, dominant allele was responsible for the identification of the gene. Furthermore, the GO ID and function of gene was different, recessive allele was associated with the MF of ubiquitin protein ligase binding (GO:0031625), while the dominant allele was involved in BP- SCF complex assembly (Table 7).

Table 7. Gene ontology analysis of three common genes identified in Norwegian and Swedish samples.

Gene ID	ONTOLOGY	GO ID	GO term description	p. adjust	Association
CUL1	MF	GO:0031625	ubiquitin protein ligase binding	0.0294597491937438	Recessive
CUL1	BP	GO:0010265	SCF complex assembly	0.0495629689636819	Dominant
CUL1	BP	GO:0010265	SCF complex assembly	0.0289889408562439	Dominant without covariance
CPL1	MF	GO:0008420	RNA polymerase II CTD heptapeptide repeat phosphatase activity	0.0302215747690859	Dominant without covariance
CPL1	MF	GO:0140994	RNA polymerase II CTD heptapeptide repeat modifying activity	0.0316445907022511	Dominant without covariance
RPK2	BP	GO:0009846	pollen germination	0.0405330538716311	Dominant
RPK2	BP	GO:0009942	longitudinal axis specification	0.0495629689636819	Dominant
RPK2	BP	GO:0009942	longitudinal axis specification	0.0289889408562439	Dominant without covariance
RPK2	BP	GO:0048653	anther development	0.0289889408562439	Dominant without covariance
RPK2	BP	GO:0048443	stamen development	0.0289889408562439	Dominant without covariance
RPK2	BP	GO:0048466	androecium development	0.0289889408562439	Dominant without covariance
RPK2	BP	GO:0010073	meristem maintenance	0.0289889408562439	Dominant without covariance
RPK2	BP	GO:0048438	floral whorl development	0.0289889408562439	Dominant without covariance
RPK2	BP	GO:0000578	embryonic axis specification	0.0353167039850309	Dominant without covariance
RPK2	BP	GO:0010152	pollen maturation	0.0362841657812691	Dominant without covariance
RPK2	BP	GO:0009956	radial pattern formation	0.04531764088098	Dominant without covariance

When the total number (29) of genes identified across different association models of Norwegian trees were compared, 19 genes were determined based on the dominant allele, 2 were associated with the recessive allele, and 11 genes were linked to genotype. Likewise, comparing the genes identified for the Swedish trees based on the different alleles and genotype yields 118 genes associated with dominant allele, 23 genes based on recessive allele, and 45 genes were associated with the genotype. Moreover, the association analysis of Total (N+S) revealed 220 genes based on the dominant allele, 40 genes associated with recessive allele, and 43 genes associated with genotype (Appendix A6).

For the Norwegian trees, gene ontology analysis conducted separately for the dominant, recessive and genotype models identified 81 different GO terms, encompassing 42 terms related to BP, 31 terms related to MF, and 8 terms related to CC. Notably, the GO terms identified by dominant, recessive, and genotype were 40, 22, and 19, respectively. Among these, the *SH3P2* gene associated with the recessive allele, contributed to the identification of 22 unique GO terms (Appendix table S4). GO enrichment dot plot depicted the top 10 (P - adjusted) enriched functions for the genes associated with PH of Norwegian trees (Figure 11). This result suggested that the enriched functions associated with the PH of the trees based on the dominant allele are mostly related to Biological Processes (BP) such as pollen germination, SCF complex assembly, intracellular potassium ion homeostasis, and anther and stamen development, and enriched terms associated with the recessive allele are related to BP such as cell plate formation involved in plant type cell wall biogenesis, and plasma membrane organization, while enriched terms related to genotype are CC - plant type vacuole membrane and MF- signalling receptor activity (Figure 11).

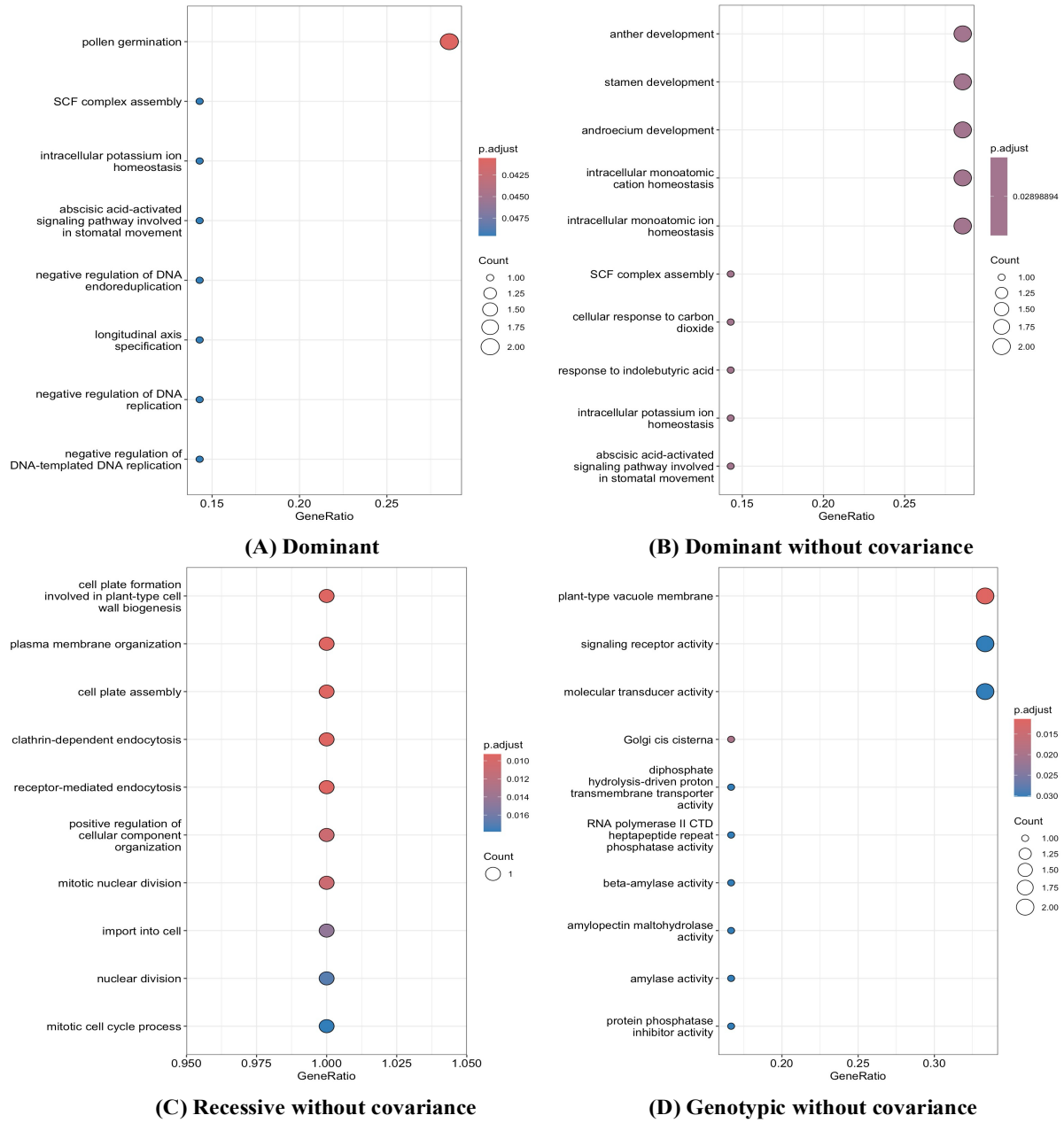


Figure 11. Dot plots showing top 10 GO terms in the GO enrichment analysis of genes identified in Norwegian trees. *(A) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant allele. (B) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant allele without covariance. (C) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with recessive allele without covariance. (D) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with genotype without covariance.*

In terms of Swedish trees, a total of 64 GO terms were identified, which includes the three different categories - BP, CC, and MF comprising of 31, 5, and 44 GO terms, respectively. Majority of GO terms (53) were identified by the genes associated with genotype and dominant alleles without covariance. Specifically, 35, 16, and 29 GO terms were identified for the dominant, recessive and genotype, respectively, with 25, 8, and 22 GO terms being unique to each corresponding allele (Appendix Table S5). The enriched function of the dominant

genes associated with PH of the Swedish trees were primarily related to molecular transducer and signalling receptor activity, detection of chemical stimulus, and plant ovule development (Appendix A7). Recessive gene were found to be mostly related with the functions such as diphosphotransferase activity, glucuronosyltransferase activity, and acyl-CoA oxidase activity. For the genotype, the enriched functions were mostly associated with calcium ion homeostasis and calcium ion transport (Appendix A 8).

Genes identified across different alleles and genotype in Total (N+S) were found to be related to 170 distinct GO term - 105 related to BP, 24 related to CC and 65 related to MF. Of these, genes associated with dominant allele accounted for 54, while those associated with the genotypic variant accounted for 38 terms. The highest number (99) of GO terms were identified by the genes associated with recessive allele. Notably, the number of unique GO terms for the dominant, recessive, and genotype allele was 33, 90 and 27, respectively. GO terms related to CC such as exoribonuclease complex (GO:1905354) and proton-transporting V-type ATPase, V0 domain (GO:0033179) were identified by the genes associated with both, dominant and recessive interaction (Appendix Table S6). It was depicted that the genes governed by dominant alleles were mostly related to gated channel activity, molecular transducer activity*, CUT* catabolic and metabolic process, ubiquitin-like protein ligase activity, and response to amino acid (Appendix A9). The enriched functions identified for the recessive genes were mostly related to anaphase promoting, nuclear ubiquitin ligase complex, *CUT* catabolic and anabolic process, histone mRNA metabolic process, fatty acid biosynthetic process, negative regulation of post-translational protein modification, and negative regulation of fatty acid metabolic process (Appendix A 10). Enriched functions identified for genotype were primarily related to Hsp90 protein binding, cyclosporin A binding, and peptidyl-prolyl cis-trans isomerase activity (Appendix A 11).

3.4.2 Gene Ontology (GO) analysis of genes associated with growth of plants under varying temperature and precipitation

Genes associated with annual growth for the different time periods of varying temperature and precipitation were identified from the SNPs that were significantly associated based on various allele and genotype models. In the LTHP/ High SPEI category during March-May, 352 genes were identified, while 118 genes were identified in the HTLP/Low SPEI category. Similarly, the period of June-August showed that 154 genes were associated with growth in the LTHP/High SPEI, whereas the genes found in the HTLP/Low SPEI category was noted as 18 (Appendix A 12). When each of the categories was assessed, it was apparent that most genes identified in the High SPEI category were associated with recessive allele in both time periods (Figure 16A). Conversely, in the Low SPEI category during March-May, genes associated with dominant allele was 86 (Figure 12 B). During the time-period of June-August, the 110 genes associated with growth in LTHP category showed a dominant allele association, while the genes associated with recessive allele was 9 in HTLP category (Figure 12 C & D).

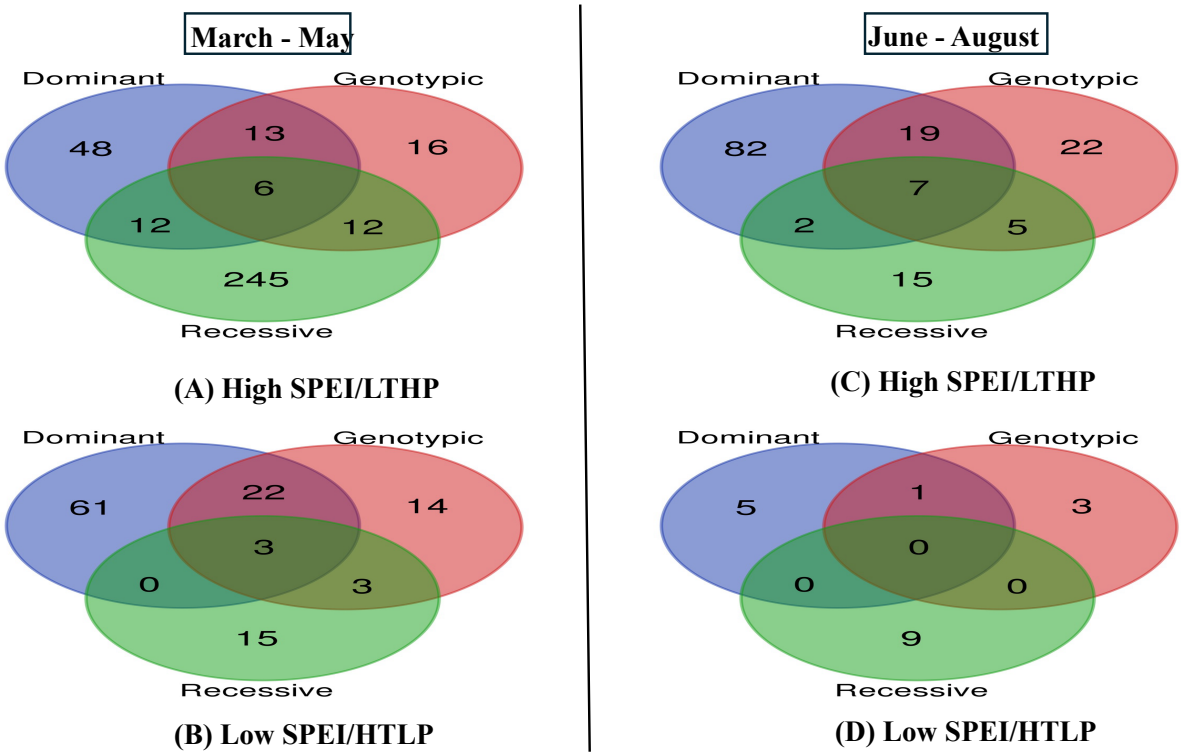
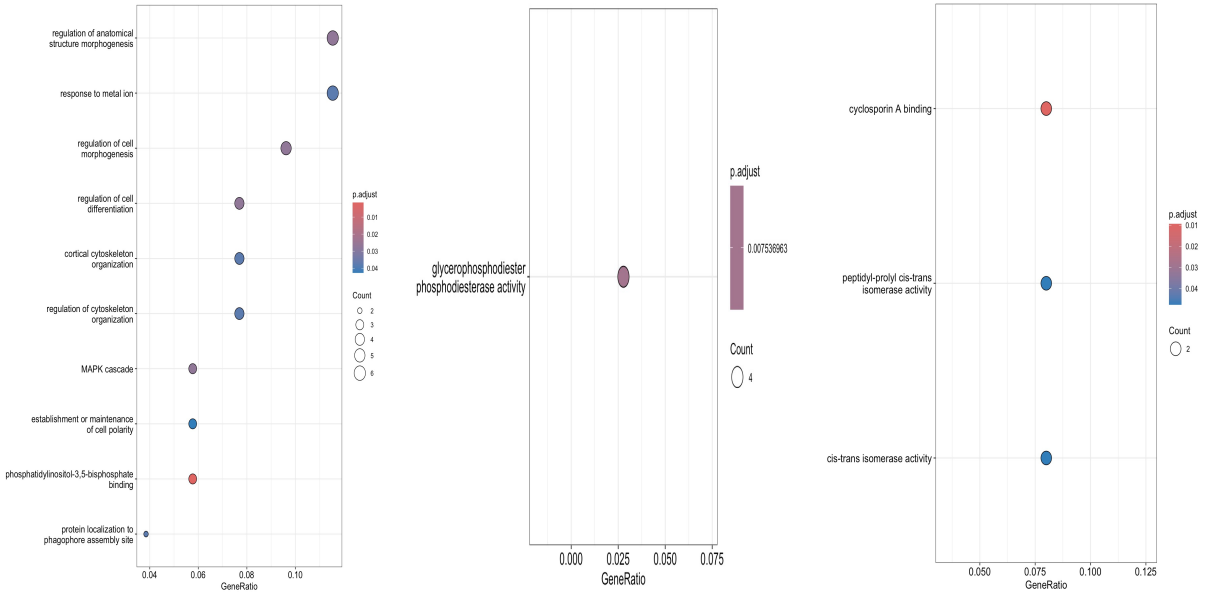


Figure 12. Venn diagram showing genes identified in different category of different time-period. *(A) Genes identified in High SPEI/LTHP of time-period March - May, (B) Genes identified in Low SPEI/HTLP of time-period March - May, (C) Genes identified in High SPEI/LTHP of time-period June -August, and (D) Genes identified Low SPEI/HTLP of time-period June - August.*

Genes associated with growth of plants with seasonal temperature and precipitation revealed different functions. During the period of March-May, a total of 95 GO terms were identified, with 17 in High SPEI category and 83 in Low SPEI category. Out of 17 GO terms in High SPEI category, 9 were related to BP and 8 were related to MF. Particularly, genes associated with dominant allele accounted for 13 GO term and genotype accounted for 3 terms, with 12 terms unique for dominant gene, whereas genotype showed no unique terms. Only 1 GO term was uniquely identified by gene associated with recessive allele (Table Appendix S4). Enriched functions for the dominant gene associated with growth in High SPEI were primarily related to BP such as regulation of anatomical and cell morphogenesis, regulation of cell differentiation, and response to metal ion (Figure 13). The sole function identified for recessive gene (ROC2/ROC1) was MF - glycerophosphodiesterase activity (Figure 13). Enriched functions for genotype were MF- peptidyl prolyl cis trans isomerase activity and cyclosporin-A binding (Figure 13). PCAP1and RAC6 genes associated with dominant allele identified 5 GO term related to BP.

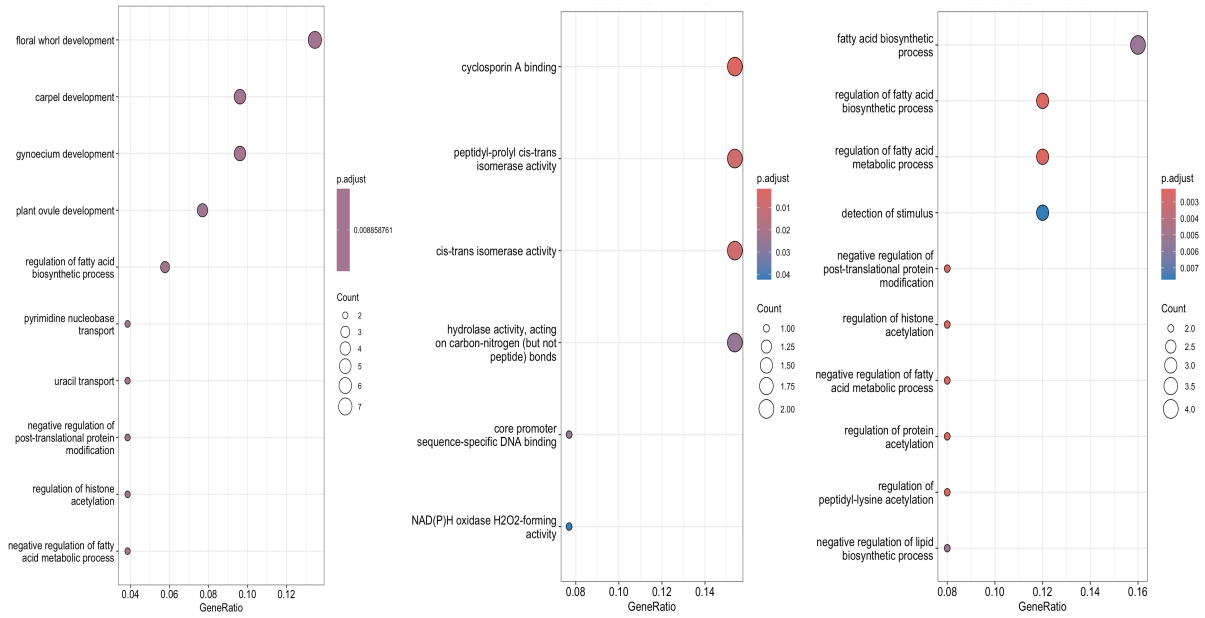
Furthermore, GO terms associated with BP, CC and MF were 73, 5, and 20, respectively in the Low SPEI category. Genes associated with dominant allele accounted for 53 terms, those associated with recessive allele for 6 terms, and those associated with the genotype for 46 terms, with 33, 3, and 25 being unique to each corresponding category (Appendix Table 4). GO ontology analysis displayed that the enriched function of the genes associated with growth of *Pinus sylvestris* based on dominant allele for High SPEI during March - May were

predominantly related to BP functions such as development of floral whorl, carpel, gynoecium, and plant ovule (Figure 14). The enriched function for genes governed by recessive allele were primarily related to MF such as cyclosporin A binding, peptidyl-prolyl cis-trans isomerase activity, and hydrolase activity acting on carbon-nitrogen (Figure 14). For the genotype association model, the enriched functions were linked to BP- regulation of fatty acid biosynthetic and metabolic process, detection of stimulus, and negative regulation of post-translational protein modification (Figure 14).



(A) Dominant without covariance **(B) Recessive without covariance** **(C) Genotypic without covariance**

Figure 13. Dot plots showing top 10 GO terms in the GO enrichment analysis of genes identified in High SPEI / LTHP category of March-May. *(A) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant allele without covariance. (B) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with recessive allele without covariance. (C) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with genotype without covariance.*



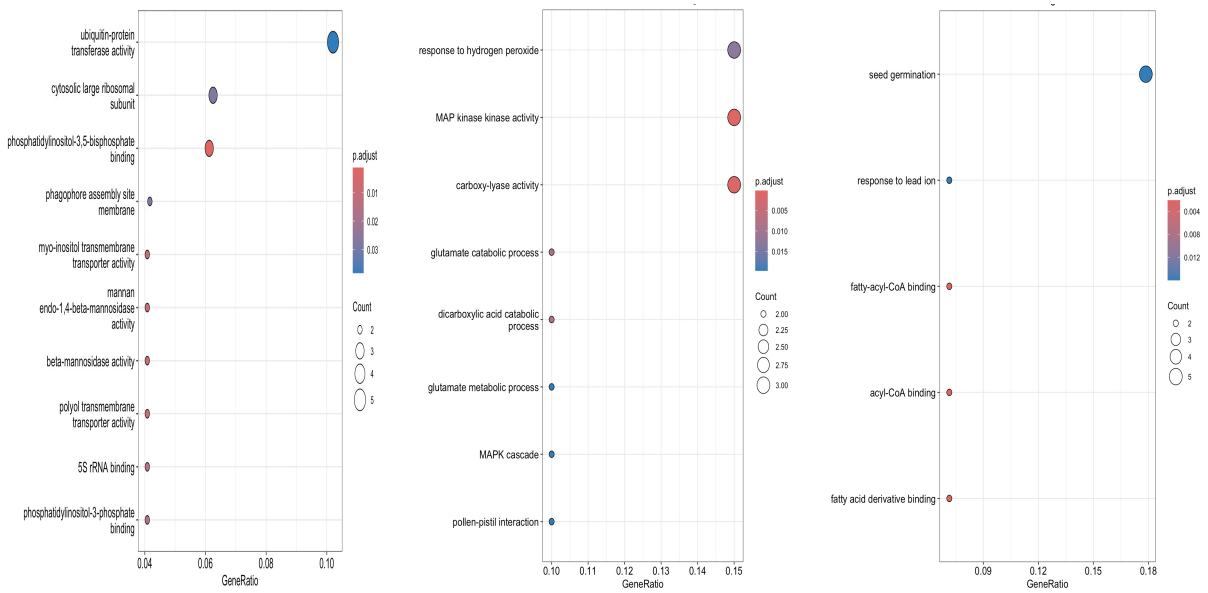
(A) Dominant without covariance **(B) Recessive without covariance** **(C) Genotypic without covariance**

Figure 14. Dot plots showing top 10 GO terms in the GO enrichment analysis of genes identified in Low SPEI / HTLP category of March-May. *(A) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant allele without covariance. (B) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with recessive allele without covariance. (C) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with genotype without covariance.*

During June-August, the total number of GO terms identified was 106, among which 24 was identified in High SPEI and 83 in Low SPEI category. In the High SPEI, GO terms assigned to BP, MF and CC were 8, 14, and 2, respectively, in which 11 were identified by genes associated with dominant allele, 8 by the recessive allele and 5 based on the genotype. Out of 10 GO terms identified for the dominant allele, only 1 was not unique, while all the terms identified for the recessive and genotype were unique (Appendix Table S5). For the genes associated with dominant allele, the enriched terms were mostly related to functions such as ubiquitin protein transferase activity, phosphatidylinositol-3,5-biphosphate binding, and phagophore assembly site membrane. The enriched function associated with growth based on recessive allele were related to response to hydrogen peroxide, MAP kinase and carboxylase activity, glutamate catabolic process and such. Moreover, enriched function associated with genotype included seed germination, response to lead ion, fatty-acyl-CoA binding and others (Figure 15).

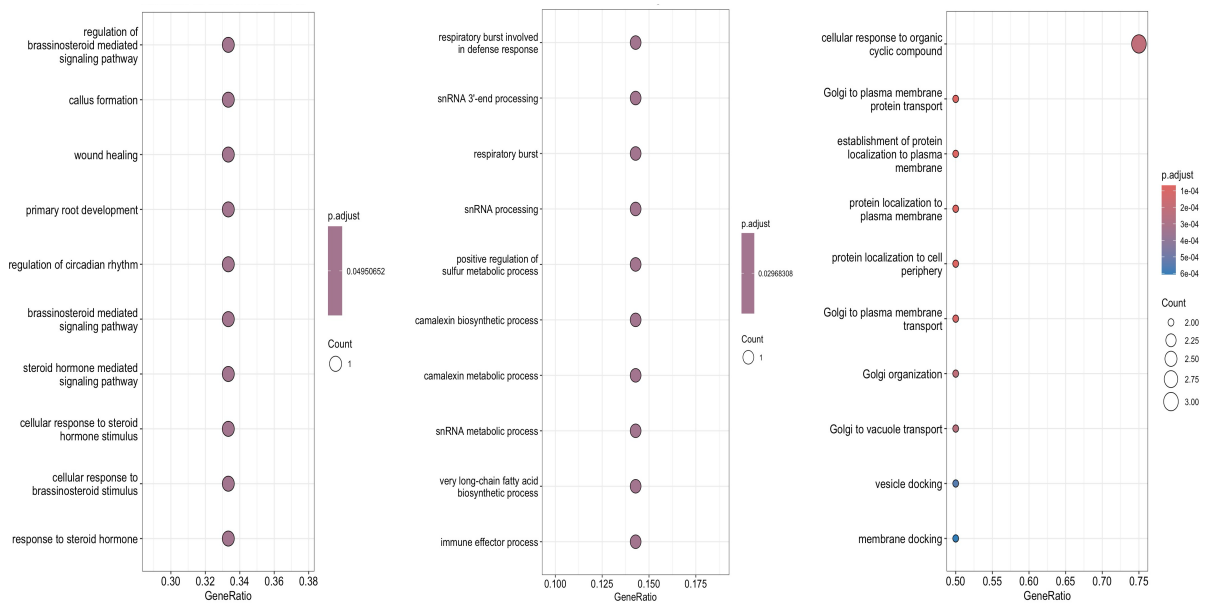
Among the 83 GO terms identified in the Low SPEI category, 11 of them were BP, 4 of them were CC, and 14 of them were MF, which were associated with genes identified by dominant allele, recessive allele and genotype. The number of GO terms identified by genes associated with dominant allele was 27 with 7 being unique. Genes associated with recessive, and genotype resulted in 26 and 50 GO terms, respectively, with 31 being unique to genotype, while all the terms identified for the recessive allele were unique (Appendix Table 5). GO dot plot indicated that the enriched function for genes associated with dominant allele in Low

SPEI category of June -August were related to regulation of brassinosteroid mediated signalling pathway, callus formation, wound healing, and primary root development, recessive gene were enriched in the functions such as respiratory burst involved in defence response, snRNA 3'-end processing, respiratory burst, and so on. For genotype associated genes, the enriched functions include cellular response to organic cyclic compound, golgi to plasma membrane protein transport, protein localization to plasma membrane, and golgi organization and transport to vacuole (Figure 16). GO analysis conducted for common genes identified in March-May and June-August showed 7 different GO terms in which 6 of them were related to MF while only 1 was related to CC (Table 8).



(A) Dominant without covariance **(B) Recessive without covariance** **(C) Genotypic without covariance**

Figure 15. Dot plots showing top 10 GO terms in the GO enrichment analysis of genes identified in High SPEI / LTHP category of June-August. *(A) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant allele without covariance. (B) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with recessive allele without covariance. (C) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with genotype without covariance.*



(A) Dominant without covariance **(B) Recessive without covariance** **(C) Genotypic without covariance**

Figure 16. Dot plots showing top 10 GO terms in the GO enrichment analysis of genes identified in Low SPEI / LTHP category of June-August. *(A) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant allele without covariance. (B) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with recessive allele without covariance. (C) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with genotype without covariance.*

Table 8. Gene ontology analysis of common genes identified in different SPEI categories during different time-period.

Genes	Ontology	GO ID	Gene Ontology term	p.adjust	SPEI / Temperature and Precipitation	Association
			March-May			
ATG18C/ATG18D/PCAP1	MF	GO:0080025	phosphatidylinositol-3,5-bisphosphate binding	0.00144477519907477	High SPEI / LTHP	DWC
MKK1/MKK2/MKK6	MF	GO:0004708	MAP kinase kinase activity	0.00405019348910528	High SPEI / LTHP	DWC
ATG18C/ATG18D	MF	GO:0032266	phosphatidylinositol-3-phosphate binding	0.0271476482980617	High SPEI / LTHP	DWC
ROC2/ROC1	MF	GO:0016018	cyclosporin A binding	0.00939780570147978	High SPEI / LTHP	RWC
ROC2/ROC1	MF	GO:0016018	cyclosporin A binding	0.00939780570147978	High SPEI / LTHP	GWC
ROC1/ROC2	MF	GO:0016018	cyclosporin A binding	0.0091407743136777	Low SPEI / HTLP	DWC
ROC1/ROC2	MF	GO:0016018	cyclosporin A binding	0.00158444578096837	Low SPEI / HTLP	RWC
ROC1/ROC2	MF	GO:0016018	cyclosporin A binding	0.011758586133132	Low SPEI / HTLP	GWC
RPL5A/RPL5B	MF	GO:0008097	5S rRNA binding	0.0091407743136777	Low SPEI / HTLP	DWC
RPL5/RPL5A/RPL5B	CC	GO:0022625	cytosolic large ribosomal subunit	0.00862574638997222	Low SPEI / HTLP	DWC
			June-August			
ATG18C/ATG18D/PCAP1	MF	GO:0080025	phosphatidylinositol-3,5-bisphosphate binding	0.00117701906024141	High SPEI or LTHP	DWC
MKK1/MKK2/MKK6	MF	GO:0004708	MAP kinase kinase activity	0.000150985480316647	High SPEI or LTHP	RWC
ATG18C/ATG18D	MF	GO:0032266	phosphatidylinositol-3-phosphate binding	0.0147061487397375	High SPEI or LTHP	DWC
ATG18C/ATG18D	CC	GO:0034045	phagophore assembly site membrane	0.027497619213474	High SPEI or LTHP	DWC
RPL5A/RPL5B	MF	GO:0008097	5S rRNA binding	0.0130574043802036	High SPEI or LTHP	DWC
RPL5/RPL5A/RPL5B	CC	GO:0022625	cytosolic large ribosomal subunit	0.027497619213474	High SPEI or LTHP	DWC
RPL34	CC	GO:0022625	cytosolic large ribosomal subunit	0.0351246995084778	Low SPEI or HTLP	DWC

3.5 DNA methylation analysis

Reduced representative bisulfite sequencing (RRBS) data obtained from the genomic DNA extracted from 30 sample trees of *Pinus sylvestris* (15 Swedish and 15 Norwegian) were analysed to investigate the epigenetic differences between Swedish and Norwegian trees, as well as to explore the potential methylation patterns associated with tree height variation.

Using the filtering criteria of adjusted p-value (meth.pval_adj), methylation difference (meth.meth_diff), and methylation mean coverage (meth.mean_cov), a total of 930 differentially methylated fragments with 5938 methylated sites were obtained. For the comparisons of NvS and TvS, the meth.meth.diff values for fragments and diff.mean.meth values for motifs were standardised for all contexts (CpG, CHG, CHH) and same values were applied for both reads (1 and 2). To reduce the list of sequences used as queries for the blast searches the standardized parameters values applied in NvS for fragments were - CpG (meth.mean.diff of +/- 15), CHG (meth.mean.diff of +/- 30), and CHH (meth.mean.diff of +/- 7) and for motifs were - CpG (diff.mean.meth of +/- 50), CHG (diff.mean.meth of +/- 50), and CHH (diff.mean.meth of +/- 50). Furthermore, the standardized parameters values applied in TvS for fragments were - CpG (meth.mean.diff of +/- 20), CHG (meth.mean.diff of +/- 30), and CHH (meth.mean.diff of +/- 7) and for motifs were - CpG (diff.mean.meth of +/- 20), CHG (diff.mean.meth of +/- 20), and CHH (diff.mean.meth of +/- 50).

Comparing the Norwegian (N) and Swedish (S), 394 differentially methylated fragments (regions) (DMRs) were identified, among which 129 were found in CpG, 128 in CHG, and 80 were found in CHH. Compared to S, 152 regions were found to be hypermethylated in N, in which 31 were identified in CpG, 66 were in CHG, and 55 in CHH. While 185 hypomethylated regions were observed which includes 98 from CpG, 62 from CHG, and 25 from CHH regions. Methylated sites found in N were 114, 177, and 868 in CpG, CHG, and CHH sites, respectively, and methylated sites of S observed in CpG was 393, CHG was 160, and CHH was 364 (Table 7). The number of DNA methylation associated motifs in NvS was highest (270) in CHH sites and lowest (138) in CHG, while 140 were identified in CpG site. There was a highest number of methylated sites observed in N in which CHH has the highest methylated sites while lowest was observed in CpG sites. In terms of S, methylation sites mainly occurred at CpG, while lowest methylated site was identified in CHG site.

With respect to Tall (T) vs Short (S), the total number of DMRs identified were 593, with 293 being identified in CpG, 188 in CHG, and 112 in CHH. When compared to Short, 292 regions were determined as hypermethylated in Tall in which, CpG, CHG, and CHH accounted for 166, 45, and 81, respectively. The total of 301 regions were found to be hypomethylated in Tall which includes 127 regions of CpG, 143 regions of CHG, as well as 31 of CHH site. Methylated sites obtained in CpG were 1609, CHG were 174, and CHH were 1270 in Tall, while methylated sites identified in Short were 441 for CpG, 591 for CHG, and 777 for CHH sites (Table 9). When DNA methylation associated motifs for TvS was observed, CHG showed the highest number of associated motifs whereas lowest number was found in CHH. The highest number of methylation site was observed in Tall which showed the highest methylated

sites in CHH, while lowest was found in CHG. Likewise, methylation sites in Short mostly occurred in CHH and lowest number of methylated sites was in CpG region. The result indicates that the methylation sites were mainly identified in CHH except for the Short, which showed the highest number methylated sites in the CpG context (Table 9).

Table 9. Differential DNA methylation in Norwegian vs. Swedish (NvS) and Tall vs. Short (TvS).

Methylated Sites	Differentially methylated fragments	Hyper methylated	Hypo-methylated	Methylated sites in N	Methylated sites in S	Methylated motifs
			N v S			
CpG	129	31	98	114	393	140
CHG	128	66	62	177	160	138
CHH	80	55	25	868	364	270
			T v S			
CpG	293	166	127	609	441	251
CHG	188	45	143	174	591	270
CHH	112	81	31	1270	777	170

DNA methylation scatterplots were created, which illustrated the relationship between DNA methylation level across different sequence contexts (CpG, CHG, and CHH) between two samples, Norwegian (N) & Swedish (S) and Tall (T) & Short (S). Each scatterplot includes Pearson correlation coefficient (r) which measures the linear relationship between methylation levels in Norwegian (N) & Swedish (S) and Tall (T) & Short (S), points observed near the diagonal indicate similar methylation levels.

The plot showed a strong positive correlation between the methylation level in N and S, high correlation (r =0.805) suggesting consistent methylation pattern between N and S in CpG site of Read 1 (Figure 28A). Similarly, strong positive correlation was observed in CHG and CHH site of Read 1 (Figure 28B and C). CpG site of Read 2 showed a very strong positive correlation indicating very high degree of similarity in methylation levels between N and S (Figure 28D). A strong positive correlation was observed in CHG site of Read 2 (Figure 28E), while CHH site of Read 2 exhibit a slightly lower but strong positive correlation which indicates some differences in methylation levels between N and S, but the trend remains positive and strong (Figure 28F). For the DNA methylation associated motif of NvS, a very strong positive correlation was observed between methylation level of N and S in all methylated sites of both, Read1 and 2, suggesting that the methylation patterns between the two samples are highly similar (Figure 29).

For Read 1 of TvS, there was a very strong positive linear relationship between T and S in CpG site (Figure 30 A), CHG site showed a strong positive linear relationship, though slightly less strong than CpG (Figure 30 B), and CHH with correlation coefficient (r =0.894) indicated a strong positive linear relationship slightly stronger than CHG site (Figure 30 C). All the sites showed strong positive relationship for Read 2 (Figure 30 E and F). All the methylated sites of both Read 1 and 2 showed the strong positive association between T and S for methylation in differentially methylated motifs (Figure 31).

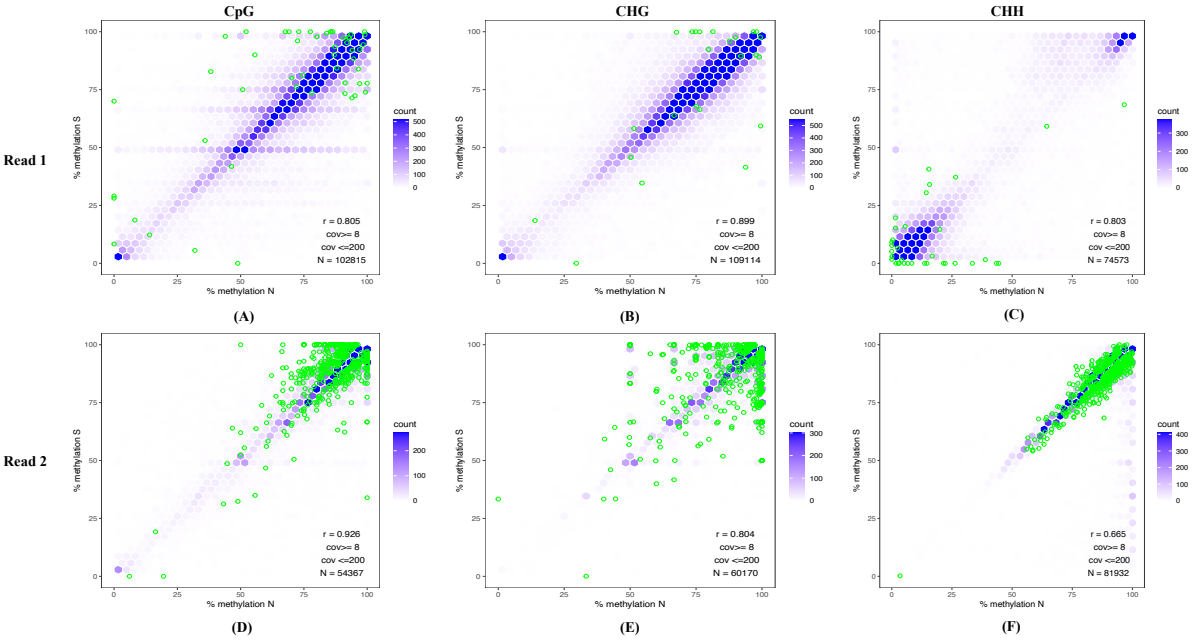


Figure 17. DNA methylation scatterplot illustrating the relationship between DNA methylation percentage of differentially methylated fragments at CpG, CHG, and CHH sites between Norwegian (x-axis) and Swedish (y axis)*. Panels A, B, and C illustrate the relationship between DNA methylation for Read 1, Panels C, D, and E represent the relationship between DNA methylation for Read 2, Blue hexagon represent high-density regions where many methylated sites have similar methylation levels in both samples. The intensity of blue color indicates count of data point, darker shades indicating higher densities, Green circles represent low-density regions with fewer data points, indicating sites with less common methylation levels between samples.*

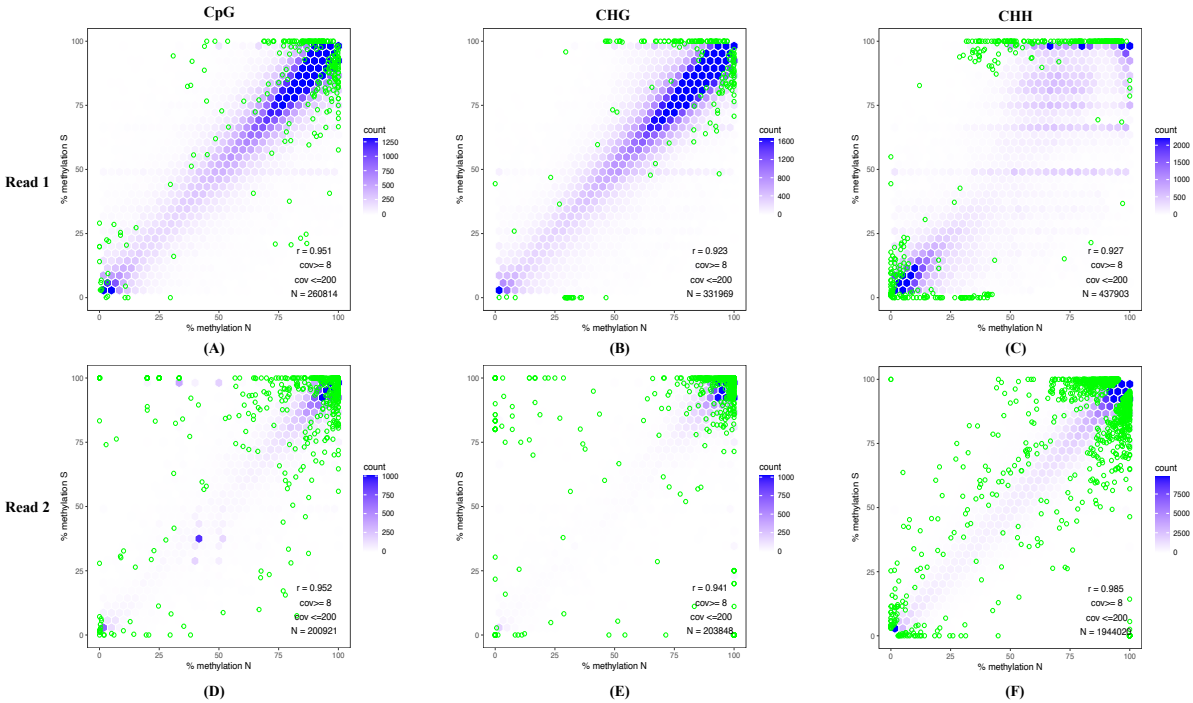


Figure 18. DNA methylation scatterplot illustrating the relationship between DNA methylation percentage for differentially methylated motifs at CpG, CHG, and CHH sites between Norwegian (x-axis) and Swedish (y axis)*. Image A, B, and C illustrates the relationship between DNA methylation for Read 1, Image C, D, and E represents the relationship between DNA methylation for Read 2, Blue hexagon represent high-density regions where many methylated sites have similar methylation levels in both samples. The intensity of blue color indicates count of data point, draker shades indicating higher densities, Green circles represent low-density region with fewer data points, indicating sites with less common methylation levels between sample.*

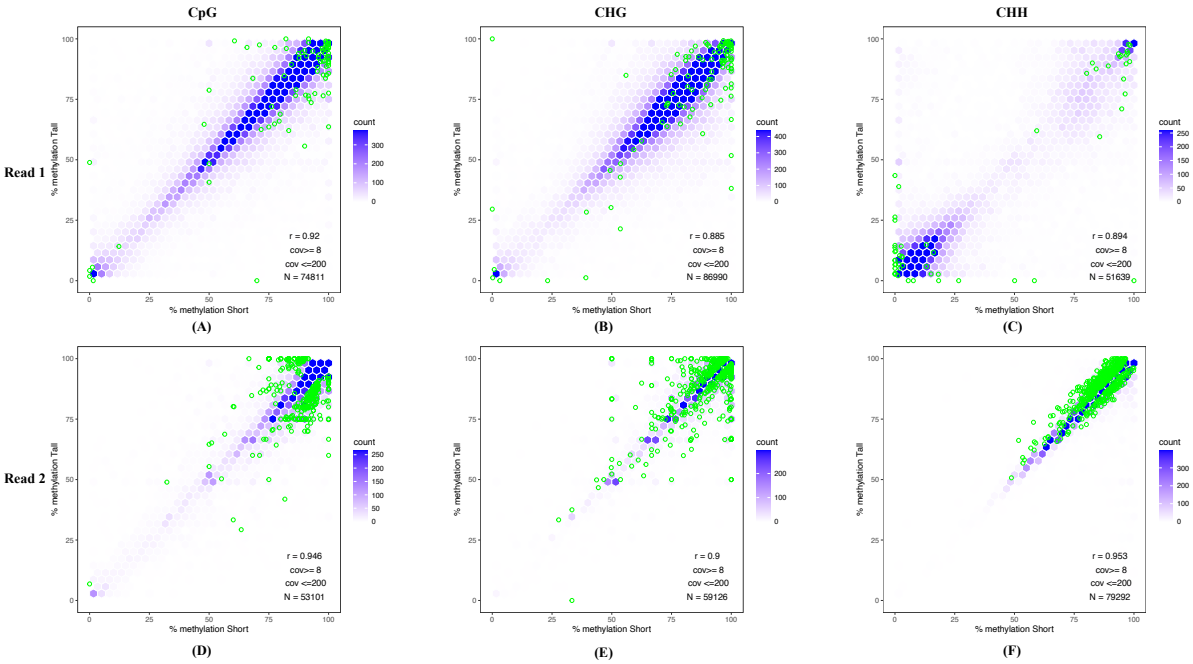


Figure 19. DNA methylation scatterplot illustrating the relationship between DNA methylation percentage for differentially methylated fragments at CpG, CHG, and CHH sites between Norwegian (x-axis) and Swedish (y axis)*. Image A, B, and C illustrates the relationship between DNA methylation for Read 1, Image C, D, and E represents the relationship between DNA methylation for Read 2, Blue hexagon represent high-density regions where many methylated sites have similar methylation levels in both samples. The intensity of blue color indicates count of data point, draker shades indicating higher densities, Green circles represent low-density region with fewer data points, indicating sites with less common methylation levels between sample.*

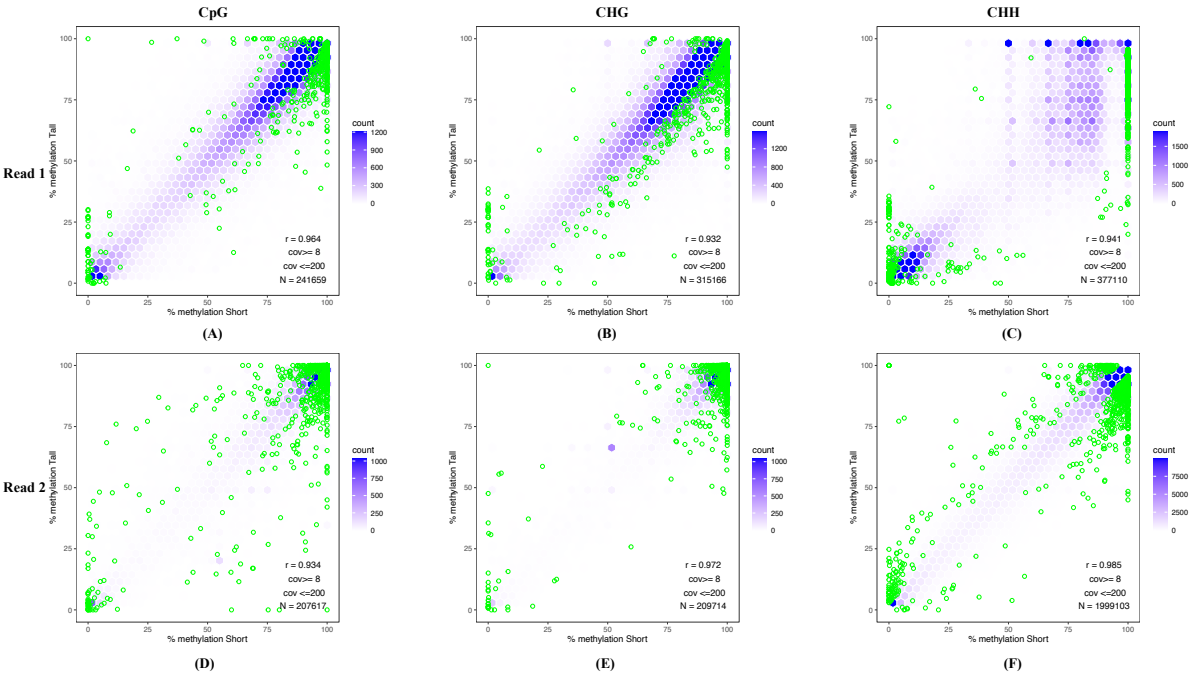


Figure 20. DNA methylation scatterplot illustrating the relationship between DNA methylation percentage for differentially methylated motifs at CpG, CHG, and CHH sites between Norwegian (x-axis) and Swedish (y axis)*. Image A, B, and C illustrates the relationship between DNA methylation for Read 1, Image C, D, and E represents the relationship between DNA methylation for Read 2, Blue hexagon represent high-density regions where many methylated sites have similar methylation levels in both samples. The intensity of blue color indicates count of data point, draker shades indicating higher densities, Green circles represent low-density region with fewer data points, indicating sites with less common methylation levels between sample.*

Principal component analysis (PCA) was performed to visualize the relationship between samples and methylated fragments creating binary variables as Norwegian vs. Swedish (NvS) and Tall vs. Short (TvS) in CpG, CHG, and CHH sites. PC1 shows the largest variation in the samples followed by other principal component in decreasing order, so the sum of all principal components is always 100%. The points contributing similar information are grouped together indicating similar DNA methylation profiles, while the points that do not cluster with others might be outliers with unique methylation profiles which can be investigated if they have any biological meaning. Points clustering near origin might indicate a homogeneous group with

similar methylation profiles or their DNA methylation profiles are not extreme in any direction captured by respective principal component. Some of the samples between groups are overlapped meaning they have shared variance or similarity in DNA methylation profiles. The first two principal components capture the most variance, but other components might also provide valuable insights and patterns not visible in first two.

For Read 1 of NvS on CpG site, PC1 and PC2 explains 8.1 and 6.4% of variance in the samples, showing dispersed samples across both axes with some observed clustering (top left). PC3 and PC4 (top right) explains 6 and 5.6% of variance in the sample capturing additional variance showing more spread-out distribution with less distinct clustering compared to PC1 vs PC2 plot, thus providing further insights into the data structure. Also, dense cluster of samples near the origin (0,0) indicated samples from multiple groups with similarity in their methylation profiles. Furthermore, for Read 2 PC1 and PC2 (bottom left) explains 12.4 and 9.2% of variance in the sample reflecting most significant differences in DNA methylation pattern and secondary pattern of variability among the samples, respectively. PC3 and PC4 (bottom right) captures 6.3 and 6% of variability in DNA methylation patterns among the samples. FURUBS45_2, FURUG25_2, and FURUG73_2 was observed as outliers and can be investigated for unique methylation patterns (Appendix A 13). Moreover, principal component analysis from Appendix figure A 14 to 17 can be interpreted and understood in a similar way.

3.6 BLAST search and analysis

Nucleic acid sequence homology searches were performed using BLASTn against database of *Pinus jeffereyi* (Accession number: SAMN12121614)for differentially methylated fragments and motifs to identify the evolutionary relationship with *Pinus jeffereyi*, a phylogenetically close relative to *P. sylvestris*. Also, BLASTx searches were performed against the RefSeqProtein database to identify differentially methylated sequences that are potentially protein coding and to predict their function or relationship with known gene sequences.

Blastx performed for differentially methylated fragments revealed significant homology with plants protein sequences, including those of *Pinus species*. Particularly, 18 differentially methylated fragments of CHG sites showed similarity with various plants proteins, with all of them were identified protein. Among them 7 of the proteins were matched with *Pinus species*, 2 were protein sequence of *Pinus koraiensis*, denoted as ORF 47a and ORF52a. Likewise, similarity with other *Pinus species* such as *Pinus ponderosa* (protein name - photosystem II protein) and *Pinus rigida/Pinus taeda* (RNA polymerase beta subunit) were identified through blast searches. For CpG sites, 6 differentially methylated fragments exhibited similarity to plant protein sequences among which 2 of them were related to (AccD) *Pinus species*, namely *Pinus densiflora var. densiflora* and *Pinus koraiensis* (ORF91). Moreover, CpG fragments were similar with disease resistance protein RGA-2 like sequence from *Coffea arabica*, while homologs in *Juglans regia* and *Cryptomeria japonica* were not identified. CHH methylation showed similarity with plant genomes however proteins remain uncharacterized (Table 11).

Methylated motif of CpG, and CHH displayed similarity with 4 and 5 plant protein sequences, respectively. In particular, 3 proteins associated with CpG motif were not identified, while a cytochrome C biogenesis protein was related to *Larix occidentalis*. CHH motifs showed similarity with 30s ribosomal protein S 12-B chloroplastic protein of *Salvia hispanica*, along with 2 uncharacterized proteins and predicted proteins from *Brassica oleracea var. oleracea*.

Among 5 similar sequences of CHG motifs, 1 remain unidentified while two showed similarities with *Pinus ponderosa* and *Pinus echinate*, name protein - photosystem II protein K and hypothetical protein NUH76_pgp002, respectively. Two of CHH motif showed similarity with plant *Cryptomeria japonica* with one having a protein named as – paired amphipathic helix protein while other was not identified (Table 12).

Table 10. Summary of total number of query and significant blast hits in BLASTn and BLASTx based on different categories for differentially methylated fragments and motifs.

BLASTn			BLASTx		
Categories	**Number of queries**	**Number of blast hits-significant E-value (< 0.01)**	**Categories**	**Number of query**	**Number of blast hits-significant E-value (< 0.01)**
Differentially methylated Fragments					
NvS - CpG (Read 1 & 2)	129	102	NvS - CpG (Read 1 & 2)	129	47
NvS - CHG (Read 1 & 2)	128	125	NvS - CHG (Read 1 & 2)	128	68
NvS CHH (Read 1 & 2)	70	65	NvS CHH (Read 1 & 2)	70	35
TvS - CpG (Read 1)	10	3	TvS - CpG (Read 1)	10	3
TvS - CpG (Read 2)	89	74	TvS - CpG (Read 2)	89	27
TvS - CHG (Read 1)	104	52	TvS - CHG (Read 1)	104	64
TvS - CHG (Read 2)	84	80	TvS – CHG (Read 2)	84	47
TvS CHH (Read 1)	55	28	TvS CHH (Read 1)	55	28
TvS CHH (Read 2)	57	55	TvS CHH (Read 2)	57	31
Differentially methylated Motifs					
NvS - CpG (Read 1 & 2)	124	93	NvS - CpG (Read 1)	28	13
NvS - CHG (Read 1 & 2)	135	127	NvS - CpG (Read 2)	96	50
NvS CHH (Read 1 & 2)	186	132	NvS - CHG (Read 1)	10	7
TvS CpG (Read 1 & 2)	177	75	NvS - CHG (Read 2)	125	77
TvS CHG (Read 1 & 2)	222	124	NvS CHH (Read 1)	46	37
TvS CHH (Read 1 & 2)	124	73	NvS CHH (Read 2)	140	67
			TvS - CpG (Read 1)	93	40
			TvS - CpG (Read 2)	84	65
			TvS - CHG (Read 1)	138	76
			TvS - CHG (Read 2)	84	65
			TvS CHH (Read 1)	49	35
			TvS CHH (Read 2)	75	48

Table 11. BLASTx of methylated fragments with plants hits at CpG, CHG, and CHH context.

Query	Number of HSPs	Lowest E-value	Accession	Description
			CpG	
dedRef_1424606-1424607_36-2	5	6,3112E-06	XP_027060706	disease resistance protein RGA2-like [*Coffea arabica]*
dedRef_571295-571296_27	5	0,047264	XP_018831472	uncharacterized protein LOC108999132 [*Juglans regia*]
dedRef_1304576-1304577_26	5	0,0233418	YP_010350660	AccD [*Pinus densiflora var. densiflora*]
dedRef_1556511-1556515_33	2	1,6489E-11	YP_001152147	ORF91 [*Pinus koraiensis*]
dedRef_930535-930536_20	5	0,00054213	YP_009528872	ATP synthase CF0 subunit I [*Dioscorea bulbifera*]
dedRef_1108957-1108958_24	5	0,318979	XP_059068972	uncharacterized protein LOC131859354 [*Cryptomeria japonica*]
			CHG	
dedRef_1761202-1761212_11	2	0,00967252	YP_001152053	ORF47a [*Pinus koraiensis*]
dedRef_778999-779000_1	5	1,9651E-06	YP_008082385	hypothetical chloroplast protein [*Pinus taeda*]
dedRef_334714-334715_38	5	6,8853E-14	XP_059067510	uncharacterized protein LOC131858325 [*Cryptomeria japonica*]
dedRef_3645364-3645365_1	5	1,2239E-07	XP_057823103	UDP-glycosyltransferase 74D1 [*Cryptomeria japonica*]
dedRef_379983-379984_40	5	0,581616	XP_050940139	transposon Tf2-1 polyprotein isoform X1 [*Cucumis melo*]
dedRef_1307669-1307670_26	5	0,026621	YP_001152109	ORF52a [*Pinus koraiensis*]
dedRef_1448271-1448272_34	5	0,00080189	YP_010309970	ribosomal protein L20 [Larix griffithii] >ref|YP_010310042.1| ribosomal protein L20 [*Larix himalaica*] >ref|YP_010310114.1| ribosomal protein L20 [*Larix kongboensis*] >ref|YP_010310186.1| ribosomal protein L20 [*Larix potaninii var. australis*] >ref|YP_010310258.1| ribosomal protein L20 [*Larix griffithii var. speciosa*]
dedRef_163909-163909_24	5	8,6795E-06	YP_009771487	photosystem I P700 apoprotein A1 [*Poecilanthe parviflora*]
dedRef_1107976-1107977_33-1	5	1,0091E-05	XP_057831826	uncharacterized protein LOC131042530 [*Cryptomeria japonica*]
dedRef_1010182-1010183_35	5	3,8973E-10	YP_008082327	RNA polymerase beta' subunit [Pinus taeda] >ref|YP_010449301.1| RNA polymerase beta' subunit [*Pinus rigida*]
dedRef_1228017-1228017_23	5	0,00010128	NP_042374	ORF58 [Pinus thunbergii]
dedRef_973251-973252_36	5	9,7209E-10	XP_020262272	uncharacterized protein LOC109838225 [*Asparagus officinalis*]
dedRef_272714-272715_22	5	1,5171E-05	YP_010526882	photosystem II protein K [Pinus ponderosa]
dedRef_536389-536389_39	5	8,6952E-05	YP_009863084	cytochrome b6 [*Anthoceros agrestis*]
dedRef_973251-973252_36	5	9,7209E-10	XP_020262272	uncharacterized protein LOC109838225 [*Asparagus officinalis*]
dedRef_272714-272715_22	5	1,5171E-05	YP_010526882	photosystem II protein K [*Pinus ponderosa*]
dedRef_536389-536389_39	5	8,6952E-05	YP_009863084	cytochrome b6 [*Anthoceros agrestis*]
dedRef_585941-585941_22	5	0,00168405	XP_060674048	photosystem II CP47 reaction center protein-like [*Ziziphus jujuba*]
			CHH	

dedRef_4150913-4150914_14	1	3,00287	XP_009111222	uncharacterized protein LOC103836680 [*Brassica rapa*]
dedRef_1703421-1703422_1	5	0,664903	XP_041183558	uncharacterized protein EDB91DRAFT_1095842 [*Suillus paluster*]

Table 12. BLASTx of methylated motifs with plants hits at CpG, CHG, and CHH context.

Query	Number of HSPs	Lowest E-value	Accession	Description
CpG				
dedRef_1855544-1855544_0	5	8,63E-11	YP_009522173	cytochrome c biogenesis protein [*Larix occidentalis]*
dedRef_290571-290573_6	1	6,12506	XP_055808402	uncharacterized protein LOC129876949 [*Solanum dulcamara*]
dedRef_290571-290573_6	1	6,12506	XP_055808402	uncharacterized protein LOC129876949 [*Solanum dulcamara*]
dedRef_987810-987811_35	1	8,75412	XP_020972789	uncharacterized protein LOC110269347 [*Arachis ipaensis*]
CHG				
dedRef_4867705-4867717_23-1	5	3,5618E-11	XP_057865118	uncharacterized protein LOC131072855 [*Cryptomeria japonica*]
dedRef_778999-779000_1	5	1,36959	YP_010449287	hypothetical protein NUH76_pgp002 [*Pinus echinata*]
dedRef_1855544-1855544_0	5	8,6309E-11	YP_009522173	cytochrome c biogenesis protein [*Larix occidentalis]*
dedRef_1186197-1186198_33	5	7,146E-13	XP_059068669	paired amphipathic helix protein Sin3-like 3 [*Cryptomeria japonica*]
dedRef_272714-272715_22	5	1,5171E-05	YP_010526882	photosystem II protein K [*Pinus ponderosa*]
CHH				
dedRef_899225-899226_33	5	7,0827E-24	XP_047953659	30S ribosomal protein S12-B, chloroplastic [*Salvia hispanica*]
dedRef_1013605-1013606_64	5	1,02227	XP_060960580	uncharacterized protein LOC115723872 isoform X2 [*Cannabis sativa*]
dedRef_4150913-4150914_14	1	3,02554	XP_009111222	uncharacterized protein LOC103836680 [*Brassica rapa*]
dedRef_1515312-1515313_33	5	6,9114E-07	XP_013615214	PREDICTED: RNA-directed DNA polymerase homolog [*Brassica oleracea var. oleracea*]

4. Discussion

The primary objective of this study was to differentiate genetically and epigenetically between Scots pine (*Pinus sylvestris)* populations originating from Swedish sowed seeds and Norwegian wind- blown seeds using SNP genotyping and DNA methylation data. Additionally, the study aimed to correlate specific SNP allele with lifetime growth rate and annual growth rate under varying temperature and water stress conditions, and to characterize the associated genes through Gene Ontology (GO) analysis.

4.1 Genotypes and SNPs categorization

In this study, genotypes for 47712 SNPs markers were initially obtained for each of 300 individual samples of *Pinus sylvestris* using the Affymetrix platform. After stringent quality control measures, only 42920 SNPs falling into Poly High Resolution (PHR) and No Minor Homozygote (NMH) categories were selected for subsequent analysis. PHR are successful SNP loci due to their reliable genotyping into well-defined clusters, polymorphic nature, and are characterized as the diploid genotypes AA, AB and BB. SNPs falling into the MonoHighResolution class exhibit a monomorphic state, lacking allelic variability and could originate from either sequencing errors within transcriptome data or actual SNPs initially identified in the SNP discovery population. Likewise, the "Other" class of SNPs includes SNPs located in the flanked region and those arising from inadequate thermodynamic properties of probes leading to non-target hybridization. Additionally, the Off-target variants (OTV) comprises SNPs that are generated by mismatches between array probes and the target DNA sequence resulting in clusters with low intensity (Howe et al., 2020).

Remarkably, a significant proportion (97.8%) of SNPs were in Hardy Weinberg Equilibrium (HWE) exhibiting a P-value of ≥ 0.01, which emphasizes the robustness of the dataset and confirms the suitability of the chosen markers for further analysis. Furthermore, 41974 SNPs demonstrated a Minor Allele Frequency (MAF) of ≥ 0.05, indicating their relevance and potential significance in genetic association studies. The higher percentage of SNPs in HWE and with MAF shows a significant level of genetic diversity within the Pinus sylvestris population. Such genetic diversity is essential for adaptability and long- term survival of the species which enables responses to environmental changes and resistance to disease.

In the majority of SNPs, observed heterozygosity (O(HET)) was higher compared to the expected heterozygosity (E(HET)), which suggests a high level of genetic diversity in a population. This higher level of genetic variability might be due to several factors such as cross fertilization between different plants which occurs through wind or insect pollination and increases likelihood of heterozygosity. Other reasons could be natural selection that maintains the genetic diversity in the population by selecting multiple alleles which result in heterozygosity in the population.

4.2 Population structure and admixture analysis

To elucidate the population structure and admixture patterns in *Pinus sylvestris*, 42,920 SNPs were analysed using STRUCTURE v.2.3.4, which utilizes an admixture model to estimate the true number of populations (K) by analysing allele frequency divergence, mean FST, and expected heterozygosity among individuals. Using the admixture model in STRUCTURE facilitated estimation of the most likely number of population clusters by maximizing the likelihood of the observed data. STRUCTURE HARVESTER processed the output results from STRUCTURE software by implementing "Evanno method" to determine the optimal K-value, which was 8 for this sample set. Admixture plots generated using fastSTRUCTURE and STRUCTURE provided clear visualization of the population structure at different K-values, ultimately confirming the presence of eight distinct clusters (K = 8) within population of *Pinus sylvestris*. These clusters represent the true underlying population structure of the sample set, which is important for understanding the evolutionary dynamics and for executing effective conservation strategies. The analysis revealed significant variation in the proportion of membership of samples in each cluster, cluster 7 showed the lowest membership proportion of 0.021, representing a minor cluster which possibly consist of unique or rare genetic subset within population. While cluster 4 indicates a major cluster with highest proportion (0.759) of membership. This emphasizes the heterogeneous nature of genetic distribution within Pinus sylvestris, which reflect a varying degree of genetic drift, gene flow, and selection pressures across different populations.

Furthermore, differences in genetic diversity among the clusters was observed, with cluster 8 showed the least (0.1836) while cluster 4 displayed the highest (0.3191), which suggest the presence of more genetically similar individuals in cluster 8 whereas individuals in cluster 4 are genetically more diverse. Conversely, genetic diversity ranging from 0.0433 to 0.580 was found between the population of *Pinus sylvestris* in which, highest diversity was noted between two most distant population (Sheller et al., 2023). The reasons for such higher genetic diversity as compared to our study might result from higher geographical distance between those populations. Geographically distant population are more likely to experience genetic drift and local adaptation, and lower chances of genes exchange which lead to greater genetic diversity within population. In contrast, populations in our study are closer geographically thus, resulting in higher gene flow and less genetic diversity.

Low nucleotide distance observed between cluster 5 and 8 (0.02999) indicates the closer relationship among the individual within these clusters. Our study also examined the mean FsT value, which measures the genetic differentiation among clusters. Highest FsT value observed in cluster 8 (0.5253) indicates significant genetic differentiation which could be due to reduced gene flow or historical isolation of this population. This suggests limited genetic exchange of this cluster with other clusters possibly due to geographical, ecological, or reproductive barriers, and such differentiation can lead to the development of unique characteristics within cluster. Conversely, low FsT value in cluster 4 (0.0118) represent a greater degree of genetic homogeneity which likely results from higher level of gene flow or a larger effective population size, enabling more frequent genetic exchanges within the cluster.

Consequently, less genetic differentiation and more uniform genetic structure in cluster 4 indicates an interconnected population. Understanding such variation in FsT values within *Pinus sylvestris* population helps in creating effective conservation strategies such as preserving the genetic uniqueness of highly differentiated clusters like cluster 8, which can maintain the genetic diversity of the species. Another study conducted in the populations of Pinus sylvestris showed genetic differentiation (Fst = 0.097) between two most distant populations of *Pinus sylvestris* (Sheller et al., 2023), which was much lower as compared to our observation. Such lower genetic differentiation suggests more uniform genetic structure potentially influenced by several factors such as geographical distance and isolation, pressure due to varying environment, historical factors like migration, and human impact such as logging, and forest management. Interestingly, the geographical distance between the population in the study was higher as compared to ours. However, there might be other reasons for lower FsT value rather than geographical distance which includes wind pollination and high rate of outcrossing among population, absence of geographical barriers allowing free gene flow, sample size as well as the number of individuals within each population. Another reason could be variations in the molecular markers used and the statistical method implemented in the calculation of FsT. Moreover, differences in environmental conditions such as climate and soil type induce varying selective pressures.

4.3 Genome Wide Association Studies (GWAS)

Single nucleotide polymorphisms (SNPs) are considered as the preferred molecular genetic markers for association studies due to their prevalence as the most abundant polymorphisms within the genome. They are distributed across both genic and intergenic regions and have capacity to denote mutational events which may induce changes in phenotypic (Cuervo-Alarcon et al., 2021). Genome Wide Association Studies (GWAS) were conducted to analyse the association of single nucleotide polymorphism (SNPs) and plant height (PH) in Scots pine samples representing Norwegian and Swedish trees as well as samples combined (Total (N+S)). The accurate identification of genetic relationships and low error rates provided by PiSy50k emphasize the suitability of using SNP array in Genome Wide Association Studies (Kastally et al., 2022). However, as large part of RNA transcriptomic was used to generate PiSy50 k SNP array, there might be some biased result due to possibilities of higher number of contig present in RNA transcriptomic than actual gene therefore, it may cause problems in downstream analyses and result in biased result (Ojeda et al., 2019).

By examining the data from 4 forest plots- Brannflata Midt (BM), Brannflata Sør (BS). Gratvikskogen (G), and Storvelta (S), this study identified significant SNPs correlated with PH to understand the genetic basis of this complex trait, which is critical for effective forest management and breeding programs. The genetic architecture of traits in forest trees is typically polygenic, which is influenced by numerous loci with minor to moderate effects. A prevalent method for identifying the genetic polymorphisms associated to adaptive traits is by testing the association between phenotypic traits and genetic variation (Cuervo-Alarcon et al., 2021).

Our findings provide valuable insights into the complex genetic interaction influencing this trait. In this study, we identified significantly associated SNPs with PH where each significant marker exhibited varying patterns of association across different alleles and origin (Norwegian or Swedish). Genome wide association studies was conducted by several studies in *Pinus species,* other forest trees and crop plants for various traits. A study conducted in three different Pine species (*Pinus sylvestris*, *Pinus uncinata*, and *Pinus mungo*) identified 118 SNP markers significantly associated with growth and phenology (Perry et al., 2022b). GWAS conducted for tree height in poplar trees identifies 41 significant SNPs associated with height (Chen et al., 2021). In Rice, GWAS conducted for plant height identified 13 associated loci for rice growing under normal water conditions, 13 associated loci for drought conditions, and 8 associated loci were identified in both conditions (X. Ma et al., 2016).

We identified 206, 987, and 1864 SNPs showing significant association with PH in Norwegian, Swedish, and combined sample Total (N+S), respectively, highlighting genetic variation influencing PH across different sample groups. The greater number of significantly associated SNPs identified in the Swedish and combined group suggests either broader genetic diversity or increased detection power due to large number of samples in these groups compared to Norwegian. Particularly, the overlap of 7 SNPs between Norwegian and Swedish groups accentuates the presence of limited shared genetic factors affecting PH among these sample groups. Norwegian trees are wind-blown seeds and might have undergone natural selection and adaptation to their environment. The differences in origin of trees likely contributes to the observed variations in SNPs association with PH in each population.

The differentiation of SNPs based on the dominant and recessive allele models, both with and without covariance enhances our understanding of the genetic contribution to PH in population of Scots pine. In the Norwegian sample, the number of significant SNPs associated with dominant allele were significantly higher as compared to associated with recessive allele in both, with covariance (53 dominant, 1 recessive) and without covariance (84 dominant, 19 recessive) association. This trend was consistent in the Swedish sample, where dominant allele associated SNPs were more common (217 dominant, 32 recessive with covariance; 279 dominant, 94 recessive without covariance). The combined analysis (Total (N+S)) supports these findings by further showing the predominant role of dominant alleles in influencing PH (344 dominant, 56 recessive with covariance; 313 dominant, 131 recessive without covariance). Our study suggests that a more detailed genetic basis for PH was observed in Swedish trees as compared to Norwegian, because the Norwegian samples size is too small (too few numbers of tall vs. short trees) to generate the same numbers of significantly associated SNPs as for the larger Swedish group.

Additionally, allele (or genotype) model*covariant interaction identified significant SNPs associated with PH. In Norwegian trees, dominant allele model*covariant interaction did not identify significant SNPs, whereas Swedish population exhibited significant interaction effects for dominant (98 SNPs), and recessive (2 SNPs) model. Combined population identified a greater number of significant SNPs under interaction effects for the dominant allele (585 SNPs), suggesting complex genetic interactions affecting PH. The unique and overlapping

SNPs identified in interaction effects (585 SNPs for dominant allele model in combined population) suggest that PH is not just influenced by individual SNPs but also by their interactions. This is particularly observed in large number of SNPs identified in dominant interaction which shows epistasis interactions where one gene's impact depends upon another gene.

Significant associations were detected in specific forest plots, indicating forest-specific genetic influences on PH. For instance, BS and S exhibited distinct sets of associated SNPs compared to G, which suggests unique genetic factors (specific SNP allele variants) having effect in these locations. There were no significant SNPs identified associated with PH in forest BM. Even the forest plot belongs to same geographical location, there could be several factors contributing to the differences in significant associations such as microclimatic variation, historical land use, sampling variation and others. The fire clearance in BM and BS plots and presence of Norwegian seeds in S and G plots represents the different management practices, fire clearance can reset the genetic composition by removing certain species or genotypes which might lead to the appearance of unique genetic factors influencing PH. Similarly, presence of Norwegian seed in the plot might introduce novel genetic variants into the population, which can cause genetic divergence. Despite belonging to the same geographical location, forest plot might experience microclimatic differences in factors such as soil composition, moisture level, and other.

Furthermore, including covariance association in the analysis provided additional insights. The number of significant SNPs associated with dominant allele and recessive alleles changes when covariance is considered versus when it was not considered, which indicates that some of the associations are stronger and more reliable when covariance is included, likely reducing false positives. Without covariance, common SNPs were found between dominant and genotype models (74 SNPs), and between recessive and genotype (27 SNPs), which highlights the complex genetic makeup of PH. Increase in the number of significant SNPs without including covariance suggests that some of the associations might be due to differences in population from different stands.

The presence of significant associations in multiple SNP markers, along with significant p-values determines the true associations as described by (Upadhyaya et al., 2013). In our study, SNPs markers identified as significantly associated in most association models (dominant, recessive and genotype) demonstrated association by having more than one SNP marker associated with plant height. However, this was not the case for the recessive model in Norwegian trees and the recessive model * covariate interaction model in Swedish trees. These findings suggest that SNP markers for plant height in Norwegian and Swedish trees are not significantly associated with plant height by recessive allele and recessive interaction models, respectively.

The varying degree of overlap among SNPs associated with different alleles model and interactions (dominant, recessive, and genotype) suggests that PH is controlled by a network of genes with both additive and non-additive effects and thus emphasizes the importance of considering both individual SNP effects and their interaction to fully understand the genetic

architecture of complex traits like plant height in Scots pine. Identification of significant numbers of SNPs across different models and populations represent the complexity of the genetic basis of PH. The limited number of SNPs shared between Norwegian and Swedish trees suggests that there are unique genetic factors involved within each population. Our findings could contribute valuable knowledge to the field of plant genetics and offer a basis for future research aimed at enhancing our understanding of this complex trait in forest trees.

Furthermore, in this study, we examined the association between SNP markers and growth of plants under varying temperature and / or water stress conditions. Growth is an important trait for examining plant responses to varying stresses, which is influenced by environmental factors (Cuervo-Alarcon et al., 2021). We implemented Standardized Precipitation Evapotranspiration Index (SPEI) to create different category (High and Low SPEI) depending on varying temperature and water stress condition. Through combining precipitation and temperature data, SPEI calculates drought index using monthly (or weekly) differences between precipitation and Potential Evapotranspiration (PET) that represent a simple climatic water balance , and enables assessment of drought severity across different period and locations as it can computed across various climatic conditions (Vicente-Serrano et al., 2010b).

By implementing SPEI, different categories (High and Low SPEI) were created based on contrasting temperature and water stress conditions. Our findings revealed a remarkable difference in the number of significantly associated SNPs across different SPEI categories and alleles. During the March-May period, when environmental stress is typically high due to low precipitation and elevated temperatures, the dominant allele association model showed the highest number of significantly associated SNPs in the Low SPEI category, which suggests that certain genetic variants might provide advantages to deal with limited water resources. In contrast, in High SPEI category, indicative of favourable temperature and precipitation, recessive allele model genotypic markers displayed a greater association with plant growth suggesting that different genetic variants might be advantageous in environments with adequate amounts of precipitation or with favourable temperature.

Furthermore, genotypic associations provided depth to the analysis by disclosing detailed interaction between genetic factors and varying temperature conditions. Particularly, significant numbers of SNPs were identified in both, Low and High SPEI categories which emphasize the complex nature of genotype-environment interactions in shaping plant growth responses. This complex nature highlights the importance of considering the combined effects of multiple genetic variants in understanding the response of plants to with fluctuation in temperature and precipitation.

During the growth period of June-August, both SPEI categories showed a greater number of significant SNPs associated with dominant allele compared to recessive allele, which suggests that certain genetic variants might play a crucial role in enabling plants to manage with various stressors during this growth period. The observed differences in significantly associated SNPs between Low and High SPEI periods likely reflect the activation of specific stress response pathways which varies with changing temperature and precipitation. During period of water

stress, plants may prioritize the expression of certain gene associated with drought conditions which results in higher number of significant SNPs associations in the dominant allele model.

Different seasonal fluctuation in temperature and precipitation imposes selective pressure on plants and affect their growth, leading to changes in adaptation, survival and thereby genetic differences. The observed differences in the number of significantly associated SNPs between different time periods (March-May and June-August) might reflect seasonal changes in light intensity, photoperiod and selective pressures. For instance, higher number of significant SNPs associated with the dominant allele model during June-Aug period might explains adaptations to warmer temperatures and increased availability of water during the season. Our findings highlight the complex nature of genetic basis of plant growth traits and explain the importance of considering both genetic and factors like temperature and precipitation in understanding plant adaptation to changing climatic conditions. Identifying specific SNPs markers associated with plant growth across diverse climatic conditions provides valuable insights into plant breeding strategies aimed at enhancing flexibility to climate change.

4.4 Gene Ontology (GO) Analysis

With an aim to identify and understand the function of the genes in which SNP markers identified in GWAS of plant height (PH) reside, Gene Ontology (GO) analysis was performed. Genes identified in the GWAS of PH within Norwegian, Swedish, and combined - Total (N+S) showed varying number of associations with SNP allele variants following dominant allele, recessive allele, and genotype models. Particularly, Swedish trees displayed a higher number of identified genes compared to Norwegian individuals.

In Norwegian trees, a greater number of genes were associated with dominant alleles as compared to the recessive allele, implying an important role of these alleles in phenotype determination. GO analysis further explains the functional importance of these alleles, which identified enriched functions primarily related to processes such as pollen germination, SCF complex assembly, and anther development. Conversely, genes associated with recessive alleles showed relation with distinct functions such as cell plate formation and plasma membrane organization, which suggests potential regulatory mechanism unique to this allele. Enriched functions associated with genotypes were related to the regulatory processes such as calcium ion homeostasis, highlighting the significance of genes for the growth and development of plants.

Similar patterns were observed in Swedish trees, which emphasize the crucial role of genotype associated genes for phenotype determination. Dominant genes were enriched in functions related to signalling receptor activity and chemical stimulus detection, which indicates the involvement of gene in environmental responsiveness. Moreover, recessive genes were mostly related to metabolic activities, implying the crucial role of genes in growth, development through involvement in conversion of nutrients, and energy.

Combined population of Norwegian and Swedish trees revealed a broad view of genetic pathways underlying PH. Shared and unique GO terms among different alleles and genotypes draw a special attention to the complexity of gene regulation and its effect on phenotype variation. Specifically, genes associated with both dominant and recessive alleles showed dual roles in cellular processes such as CUT catabolic and metabolic processes, indicating the multiple roles of those genes. Our study provides valuable insights into the complex interaction between genetic variation and phenotype determination in trees.

Common genes identified in Norwegian and Swedish are related to various functions, such as *CUL1* gene is related to both Molecular function (ubiquitin protein ligase binding) and biological processes (SCF complex assembly). *CUL1* shows continual expression across entire life cycle of plants, from the inception of zygote to the production of pollen. The essential role of *CUL1* and SCF complexes in regulating auxin responses throughout various stage of growth and development has been reported in (Hellmann et al., 2003). *CPL1* gene is associated with Molecular function related to RNA polymerase II CTD heptapeptide repeat phosphatase and modifying activity. CPL1 is essential for transcriptional regulation and involved in miRNA biogenesis, plant growth and stress responses. Delayed flowering phenotype in Arabidopsis thaliana is addressed due to loss of *CPL1* function (Yuan et al., 2021). Another similar gene identified is *RPK2,* which was related to several functions related to BP such as pollen germination, longitudinal axis specification, anther, stamen, and androecium development, meristem maintenance, floral whorl development, embryonic axis specification, pollen maturation, and radial pattern formation. *RPK2* was found to be involved in regulating root growth by controlling cell proliferation and influencing meristem size in *Arabidopsis thaliana* (Racolta et al., 2018).

Moreover, Gene Ontology Analysis (GO) analysis was conducted for the genes associated with growth of plants to understand the Biological Process (BP), Cellular components (CC), and Molecular Function (MF) of the genes associated with growth of plants with seasonal temperature and precipitation. The study provided insights into the distinct significant role of various genes across different environmental conditions, which highlights the dynamic interaction between the genetic factors and environmental stimuli in influencing plant growth and development by identifying Gene Ontology (GO) terms associated with the High SPEI and low SPEI categories during two different growth period (March-May and June-August).

During the period of March-May, a total of 95 GO terms were identified with significant differences observed between High and Low SPEI category. In High SPEI, genes associated with dominant allele were related to the regulatory processes such as anatomical and cell morphogenesis, cell differentiation, and response to metal ions. Such functions are pivotal for early growth and development process in plants, which shows the essential role of dominant alleles in vigorous growth and development of plants in the favourable environmental conditions. The sole function identified for the recessive gene in High SPEI was glycerophosphodiesterase activity, which highlights the significant involvement of recessive allele in physiological process in plants, ultimately growth and development of plants.

Genotype-related functions such as peptidyl prolyl cis trans isomerase activity and cyclosporin A- binding underscore the importance of protein folding and stress response mechanism.

In contrast, Low SPEI category identified various functions associated with alleles and genotype. Genes governed by dominant allele were predominantly associated with various floral and reproductive development such as development of floral whorl, carpel, gynoecium, and plant ovule. Such reproductive functions observed under conditions of lower moisture indicates an adaptive strategy followed by plants to prioritize the reproductive success over vegetative growth in water stress. Association of recessive gene with the stress response mechanism such as cyclosporin A binding and hydrolase activity shows the essential role of recessive allele in maintaining cellular functions under unfavourable conditions or water stress condition in Low SPEI period. Genotype related functions include regulation of fatty acid biosynthesis processes and detection of stimuli indicates a wide range of adaptive responses to environmental stress.

Furthermore, 106 GO terms were identified across High SPEI and Low SPEI category during the growth period of June- August. Dominant genes in the High SPEI category were related with cellular functions such as ubiquitin protein transferase activity and phosphatidylinositol-3,5-biphosphate binding which play a crucial role in regulation of protein. The association of dominant gene with phagophore assembly site membrane suggest an active role in autophagy and cellular survival and maintenance during the period of high-water availability. Recessive genes were related with the functions such as response to hydrogen peroxide and MAP kinase activity which shows the role of recessive allele in reducing oxidative stress. The association of genotypes with the function such as seed germination and response to lead ion highlights various physiological processes in the adequate moisture conditions. In Low SPEI period of June-August, dominant genes were primarily related with brassinosteroid – mediated signalling and primary root development which highlights the role of dominant allele to manage stress in plants by controlling root growth and hormones. Genes governed by recessive allele show association with stress response function which highlight the continuity of recessive allele in helping plants to handle stressed conditions. Genotype related functions such as protein transport and golgi organization highlights the influence of genotype in molecular machinery of these process which determine ability of cell to survive in adverse environmental condition. GO enrichment analysis for different genes associated with allele and genotype elucidate the complex genetic strategies exerted in plants for growth and survival with varying temperature and water stress conditions. Dominant genes typically control broad regulatory and developmental processes, whereas recessive genes and genotype specific functions frequently contribute to specialized roles in maintaining cellular functions and stress response.

4.5 DNA methylation analysis

Analysing Reduced Representative Bisulfite Sequencing (RRBS) data obtained from 30 individual sample of Pinus sylvestris, including Norwegian and Swedish samples revealed notable epigenetic differences between these two origins. Also, we analysed the data to

examine the methylation pattern associated with tree height (Tall vs Short) and observed significant differences.

By comparing two origins of trees Norwegian and Swedish, we identified 394 differentially methylated regions (DMRs) which includes 129 in CpG, 128 in CHG, and 80 in CHH sites. Among these DMRs, 152 regions were hypermethylated while 185 regions were hypomethylated. Further, differences in the distribution of the methylated sites were observed in between Norwegian and Swedish trees. Norwegian trees showed higher methylation in CHH sites (868) while lowest in CpG sites (114). In contrast, higher methylation in Swedish trees was found in CpG site (393), whereas lower methylation was observed in CHG site (160). Such differences in methylation profiles observed among these two different origins of trees indicates the different epigenetic mechanism exerted by trees which might be related to the local adaptations or due to genetic differentiation. When considering the potential functional effects, only CpG methylations has been associated with gene expression in *Pinus taeda* (Gardner et al., 2023). Our study shows higher CpG methylation in Swedish trees as compared to Norwegian trees, which suggest that Swedish trees might engage more broadly in gene regulation through CpG methylation. Higher level might facilitate optimization of gene expression more precisely which enable them to adapt more effectively to their specific environmental conditions such as climate, soil composition and other factors. In contrast, lower CpG methylation observed in Norwegian trees indicates less dependence on CpG methylation for regulation of gene expression. In plants, CHG methylation is associated with gene splicing and modifications in methylation status have the potential to disrupt precise splicing of gens which may lead to harmful effects (Gardner et al., 2023). Lower level of CHG methylation found in Swedish trees imply decreased role of CHG methylation in splicing mechanism of trees, which might indicate an adaptation to reduce the harmful effects that could arise from epimutations in these regions.

Furthermore, in plants, CHH methylation is predominantly associated with the silencing of transposable elements (TEs) and occurrence of epimutations in these regions are generally low (Gardner et al., 2023). In Norwegian trees, higher level of CHH methylation shows a strong mechanism for maintaining genome integrity by silencing TEs. The lower CHH methylation in Swedish trees imply either lower TEs activity or evolution of alternative mechanism for managing TE activity. Such differences in CHH methylation between these origin highlights diverse strategies employed by them to maintain genomic stability.

When methylation pattern associated with tree height was compared, 593 DMRs were identified between Tall (T) and Short (S) trees, among which 293 were in CpG, 188 in CHG, and 112 in CHH site. Identification of 593 DMRs between tall and short trees highlights the regulatory function of DNA methylation in controlling processes associated with growth. Comparing Tall vs Short identified 292 hypermethylated and 301 hypomethylated regions. Moreover, observed methylation site count significantly varied between tall and short trees, tall trees showed higher methylated site in CHH (1270) while lower methylated site in CHG (107). For Short trees, higher methylated site was identified in CHH (777) and fewer methylated site was found in CpG (441). These variation in methylation pattern imply possible

epigenetic basis for variation of height in *Pinus sylvestris*, where different methylation patterns correlate with trait. Such variation in between tall and short trees might be resulted from various factors such as growth, development and adaptations to environment. Higher methylation in CHH can also be influenced by microenvironment, which includes several factors such as soil, water temperature, humidity, pest and others. Microenvironment accounted for 0.086% of phenotypic variation in Plant height has been recorded (Hu et al., 2015) . Tall trees generally experience more stress due to light exposure, temperature fluctuation, which can induce stress in plants. CHH methylation is found to be associated with the regulation of genes related to stress response (Wang et al., 2022). The greater number of CHH methylated sites in tall trees might indicate an increased need for regulation of gene expression to manage the physiological and environmental stress associated with greater height. Since CHH methylation is related to plant immune responses (D. Xiao et al., 2021), our findings suggests that both tall and short trees might have an enhanced immune responses due to their elevated CHH methylation levels. Our findings offer valuable insights into the mechanism underlying plant adaptation and defence strategies. This finding has significant implications for forestry and agricultural practices as it highlights the importance of DNA methylation patterns, particularly CHH methylation, which could provide insight into the breeding programs aimed at improving disease resistance. Understanding this pattern can help breeders to perform selective breeding with improved CHH methylation pattern, enhancing the ability of plants to withstand pathogens and environmental challenges and thus result in sustainable production. Variation in the methylation patterns observed between Swedish and Norwegian as well as between tall and short trees draw an attention to the complex nature of epigenetic regulation influenced by environmental and genetic factor.

4.6 BLAST search and analysis

BLASTn was performed against the assembled genome of *Pinus jeffreyi*, whereas BLASTx was performed against the Ref-seq protein database. Using BLASTn, nucleic acid sequence homology searches against the *Pinus jeffreyi* database (Accession number: SAMN1212614) was performed with aim to identify evolutionary relationship of differentially methylated fragments and motifs with *Pinus jeffreyi*, which is a phylogenetically close relative of Pinus sylvestris. Most of the methylated fragments and motifs at CpG, CHG, and CHH sites showed significant similarity with sequences from *P. jeffreyi*, representing the genetic relatedness and conservation of methylation patterns within the species. E-value quantifies the statistical significance of similarity search in BLAST, where lower E-value suggests more significant sequence similarity (the better is the hit i.e., the sequences are homologs) and higher E-value suggests that the similarities between query and retrieved database sequence are by chance (Kerfeld & Scott, 2011)

BlASTX searches against the RefseqPROTEIN database aimed to identify protein coding potential of differentially methylated sequences and to predict functions or relationship with known gene sequences. From evolutionary perspective, protein sequences are more conserved as compared to nucleotide sequences and also the detected protein hits might have been annotated facilitating the observation of protein functions (Bioinformatics Software | QIAGEN Digital Insights). The result showed significant homology with plant protein sequences especially those from Pinus species, indicates evolutionary conservation and potential functional roles of these methylated regions.

For the CHG context, 18 differentially methylated fragments showed similarity with plant proteins, which suggests conservation of these sequences across plant genomes. Particularly, 7 fragments matched with *Pinus species*, and 2 were identified in *Pinus koraiensis* (ORF 47a and ORF52a). Additionally, other matches include photosystem II protein from *Pinus ponderosa* and RNA polymerase beta subunit from *Pinus rigida*/*Pinus taeda*e. Photosystem II (PS II) is a multi-pigment protein complex located within the thylakoid membranes of oxygenic photosynthetic organisms, such as cyanobacteria, algae, and plants. It plays a crucial role in splitting water, reducing plastoquinone, and oxygen evolution (Lu, 2016). These matches suggests that CHG methylation sites in *Pinus sylvestris* are involved in important biological processes such as photosynthesis and transcription and are highly conserved among *Pinus species.* In CpG, 6 differentially methylated fragments showed similarity to plant protein sequences, with 2 were related to *Pinus species*, namely *Pinus densiflora var. densiflora* and *Pinus koraiensis* (ORF91). Furthermore, 1 CpG fragment exhibiting similarity with a disease resistance protein RGA-2 like sequence from *Coffea arabica* suggests potential role in plant defense mechanisms. RGA2 gene was identified playing an essential role in disease resistance against *Phytophthora capsici* Leonian in *Capsicum annum* (Y.-L. Zhang et al., 2013). Although CHH methylation sites showed similarity across plant genomes, protein remains uncharacterized (there was a hit, but protein has no identity or function) which indicates the presence of potential novel proteins or genomic regions with unidentified functions and requires further study.

BLASTx search of methylated motifs (CpG, CHG, and CHH) revealed several plant protein sequences which elucidate the conservation and possible functional roles of these motifs. CpG motifs showed identification with 4 plant protein sequences including a cytochrome C biogenesis protein from *Larix occidentalis*, which is essential for electron transport in photosynthesis. Cytochrome c proteins are essential heme carriers, which undergo a complex

post-translational biosynthesis known as cytochrome c biogenesis, and are found in most living organisms and are crucial for energy production and various cellular processes, (Verissimo & Daldal, 2014). The identification of uncharacterized (there was a hit, but protein has no identity or function) proteins suggests need for further research to identify their roles. In CHH motifs 5 plant protein sequences were identified, a 30s ribosomal protein S 12-B chloroplastic protein identified from *Salvia hispanica*, which is crucial for protein synthesis in chloroplasts. Ribosomal protein S12 is located at the interface between ribosomal subunits, where it is play an essential roles to interact with tRNA molecules and large subunit (Cukras et al., 2003). The resemblance to predicted proteins from *Brassica oleracea var. oleracea* and 2 uncharacterized proteins suggests the presence of conserved but functionally obscure regions in plant genomes.

CHG motifs showed similarity with 5 plant sequences, 2 were related to *Pinus ponderosa* and *Pinus echinate*, with protein sequence- a photosystem II protein and a hypothetical protein NUH76_pg002, respectively. These proteins are likely involved in crucial biological processes such as photosynthesis. Intrinsic and extrinsic proteins of Photosystem II helps to neutralize light stress in plants (Landi & Guidi, 2023). Moreover, another CHH motif showed similarity with a paired amphipathic helix protein from *Cryptomeria japonica*, which is potentially involved in transcriptional regulation. The significant homology of differentially methylated fragments with proteins of *Pinus species* and other plant species emphasizes evolutionary conservation of these methylated patterns. Identification of proteins involved in photosynthesis, transcription, and disease resistance suggests that methylation might play a key role in regulation of essential physiological processes in *Pinus sylvestris*. Additionally, presence of uncharacterized and hypothetical proteins shows that there are still unknown aspects of plant genomics and epigenomics which require further exploration. To recapitulate, BLASTn and BLASTx searches and analyses provided a detailed overview of evolutionary relationships and potential functions of differentially methylated sequences in P. sylvestris with homology to *Pinus jeffreyi* and other plant species. These findings helped to characterise the methylated sequences in *Pinus sylvestris*, find a homologous sequences in other species and their potential roles in important biological functions, thereby facilitating future investigations to completely understand the functional importance of DNA methylation in *Pinus sylvestris*.

5. Conclusion

A detailed analysis of genetic and epigenetic factors influencing the traits of *Pinus sylvestris*, particularly plant height and annual growth unveils the intricate relationship between genomic variations, temperature and precipitation, and phenotypes. Firstly, population structure analysis of 310 individuals of Scots pine using SNPs markers highlights the presence of 8 distinct clusters (K = 8) within the sampled trees, with significant genetic diversity observed across different forest plots and between Norwegian and Swedish trees. And the largest cluster 4 most likely represented trees originating from Swedish seed.

Moreover, Genome Wide Association Studies (GWAS) offers valuable insights into the genetic framework of plant height and growth by revealing significant SNPs markers associated with these traits across different sample groups and climatic conditions. The predominance of dominant alleles found to correlate with plant height suggests their significant role in determining phenotypic variations, thereby holding potential implications for plant breeding and selection strategies. Notable differences in DNA methylation patterns were identified between Swedish and Norwegian trees as well as between tall and short individuals. These epigenetic variations might reflect the different environmental life histories between seeds of different geographical origin, and/or to the adaptability of traits and capacity to respond to environmental signals, thereby influencing growth and development of Scots pine trees.

Furthermore, our detailed analysis contributes to a deeper understanding of the underlying genetic and epigenetic mechanisms governing Scots pine traits. These findings provide valuable insights for conservation practices and sustainable forest management as well as breeding approaches directed towards enhancing resilience and productivity in *Pinus sylvestris*.

6. References

Alakärppä, E., Salo, H. M., Valledor, L., Cañal, M. J., Häggman, H., & Vuosku, J. (2018). Natural variation of DNA methylation and gene expression may determine local adaptations of Scots pine populations. *Journal of Experimental Botany*, *69*(21), 5293–5305.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Bartels, A., Han, Q., Nair, P., Stacey, L., Gaynier, H., Mosley, M., Huang, Q. Q., Pearson, J. K., Hsieh, T.-F., An, Y.-Q. C., & Xiao, W. (2018). Dynamic DNA Methylation in Plant Growth and Development. *International Journal of Molecular Sciences*, *19*(7), Article 7. https://doi.org/10.3390/ijms19072144

Bewick, A. J., & Schmitz, R. J. (2017). Gene body DNA methylation in plants. *Current Opinion in Plant Biology*, *36*, 103–110. https://www.sciencedirect.com/science/article/pii/S1369526616301297

Bilginer, U., Ergin, M., Demir, E., Yolcu, H. I., & Karsli, B. A. (2022). *Detection of genetic diversity in cattle by microsatellite and SNP markers-a review.* https://www.researchgate.net/profile/Eymen-Demir/publication/366696320_Detection_of_genetic_diversity_in_cattle_by_microsatellite_and_SNP_markers_-_a_review/links/63d681acc465a873a26a5602/Detection-of-genetic-diversity-in-cattle-by-microsatellite-and-SNP-markers-a-review.pdf

Botton, A., Galla, G., Conesa, A., Bachem, C., Ramina, A., & Barcaccia, G. (2008). Large-scale Gene Ontology analysis of plant transcriptome-derived sequences retrieved by

AFLP technology. *BMC Genomics*, *9*(1), 347. https://doi.org/10.1186/1471-2164-9-347

Brichta, J., Šim\uunek, V., Bílek, L., Vacek, Z., Gallo, J., Drozdowski, S., Bravo-Fernández, J. A., Mason, B., Roig Gomez, S., & Hájek, V. (2024). Effects of Climate Change on Scots Pine (Pinus sylvestris L.) Growth across Europe: Decrease of Tree-Ring Fluctuation and Amplification of Climate Stress. *Forests*, *15*(1), 91. https://www.mdpi.com/1999-4907/15/1/91

Brichta, J., Vacek, S., Vacek, Z., Cukor, J., Mikeska, M., Bílek, L., Šimůnek, V., Gallo, J., & Brabec, P. (2023). Importance and potential of Scots pine (*Pinus sylvestris* L.) in 21 st century. *Central European Forestry Journal*, *69*(1), 3–20. https://doi.org/10.2478/forj-2022-0020

Burghardt, L. T., Young, N. D., & Tiffin, P. (2017). A Guide to Genome-Wide Association Mapping in Plants. *Current Protocols in Plant Biology*, *2*(1), 22–38. https://doi.org/10.1002/cppb.20041

Carlisle, A., & Brown, A. H. F. (1968). Pinus Sylvestris L. *Journal of Ecology*, *56*(1), 269–307. https://www.jstor.org/stable/2258078

Catoni, M., Tsang, J. M., Greco, A. P., & Zabet, N. R. (2018). DMRcaller: A versatile R/Bioconductor package for detection and visualization of differentially methylated regions in CpG and non-CpG contexts. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky602

Chen, Y., Wu, H., Yang, W., Zhao, W., & Tong, C. (2021). Multivariate linear mixed model enhanced the power of identifying genome-wide association to poplar tree heights in a randomized complete block design. *G3 Genes|Genomes|Genetics*, *11*(2), jkaa053. https://doi.org/10.1093/g3journal/jkaa053

Cuervo-Alarcon, L., Arend, M., Müller, M., Sperisen, C., Finkeldey, R., & Krutovsky, K. V. (2021). A candidate gene association analysis identifies SNPs potentially involved in drought tolerance in European beech (Fagus sylvatica L.). *Scientific Reports*, *11*(1), 2386. https://doi.org/10.1038/s41598-021-81594-w

Cukras, A. R., Southworth, D. R., Brunelle, J. L., Culver, G. M., & Green, R. (2003). Ribosomal Proteins S12 and S13 Function as Control Elements for Translocation of the mRNA:tRNA Complex. *Molecular Cell*, *12*(2), 321–328. https://doi.org/10.1016/S1097-2765(03)00275-2

Durrant, T. H., De Rigo, D., & Caudullo, G. (2016). Pinus sylvestris in Europe: Distribution, habitat, usage and threats. *European Atlas of Forest Tree Species*, *14*, 845–846. https://www.researchgate.net/profile/Giovanni-Caudullo/publication/299470777_Pinus_sylvestris_in_Europe_distribution_habitat_usage_and_threats/links/61ee73a8dafcdb25fd4a0b70/Pinus-sylvestris-in-Europe-distribution-habitat-usage-and-threats.pdf

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Elhamamsy, A. R. (2016). DNA methylation dynamics in plants and mammals: Overview of regulation and dysregulation. *Cell Biochemistry and Function*, *34*(5), 289–298. https://doi.org/10.1002/cbf.3183

Erdmann, R. M., & Picard, C. L. (2020). RNA-directed DNA Methylation. *PLOS Genetics*, *16*(10), e1009034. https://doi.org/10.1371/journal.pgen.1009034

Fjellstad, K. B., & Skrøppa, T. (2020). State of forest genetic resources in Norway 2020. *NIBIO Rapport*. https://nibio.brage.unit.no/nibio-xmlui/handle/11250/2720189

Foerster, A. M., & Scheid, O. M. (2010). Analysis of DNA Methylation in Plants by Bisulfite Sequencing. In I. Kovalchuk & F. J. Zemp (Eds.), *Plant Epigenetics* (Vol. 631, pp. 1–11). Humana Press. https://doi.org/10.1007/978-1-60761-646-7_1

Fraga, M. F., & Esteller, M. (2002). DNA Methylation: A Profile of Methods and Applications. *BioTechniques*, *33*(3), 632–649. https://doi.org/10.2144/02333rv01

Galdina, T., & Khazova, E. (2019). Adaptability of Pinus sylvestris L. to various environmental conditions. *IOP Conference Series: Earth and Environmental Science*, *316*(1), 012002. https://iopscience.iop.org/article/10.1088/1755-1315/316/1/012002/meta

Gallego-Bartolomé, J. (2020). DNA methylation in plants: Mechanisms and tools for targeted manipulation. *New Phytologist*, *227*(1), 38–44. https://doi.org/10.1111/nph.16529

Gardner, S. T., Bertucci, E. M., Sutton, R., Horcher, A., Aubrey, D., & Parrott, B. B. (2023). Development of DNA methylation-based epigenetic age predictors in loblolly pine (Pinus taeda). *Molecular Ecology Resources*, *23*(1), 131–144. https://doi.org/10.1111/1755-0998.13698

Gawehns, F., Postuma, M., van Antro, M., Nunn, A., Sepers, B., Fatma, S., van Gurp, T. P., Wagemaker, N. C. A. M., Mateman, A. C., Milanovic-Ivanovic, S., Groβe, I., van Oers, K., Vergeer, P., & Verhoeven, K. J. F. (2022). epiGBS2: Improvements and evaluation of highly multiplexed, epiGBS-based reduced representation bisulfite sequencing. *Molecular Ecology Resources*, *22*(5), 2087–2104. https://doi.org/10.1111/1755-0998.13597

Gruntman, E., Qi, Y., Slotkin, R. K., Roeder, T., Martienssen, R. A., & Sachidanandam, R. (2008). Kismeth: Analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics*, *9*(1), 371. https://doi.org/10.1186/1471-2105-9-371

He, X.-J., Chen, T., & Zhu, J.-K. (2011). Regulation and function of DNA methylation in plants and animals. *Cell Research*, *21*(3), 442–465. https://doi.org/10.1038/cr.2011.23

Hellmann, H., Hobbie, L., Chapman, A., Dharmasiri, S., Dharmasiri, N., del Pozo, C., Reinhardt, D., & Estelle, M. (2003). Arabidopsis AXR6 encodes CUL1 implicating SCF E3 ligases in auxin regulation of embryogenesis. *The EMBO Journal*, *22*(13), 3314–3325. https://doi.org/10.1093/emboj/cdg335

Henderson, I. R., Chan, S. R., Cao, X., Johnson, L., & Jacobsen, S. E. (2010). Accurate sodium bisulfite sequencing in plants. *Epigenetics*, *5*(1), 47–49. https://doi.org/10.4161/epi.5.1.10560

Howe, G. T., Jayawickrama, K., Kolpak, S. E., Kling, J., Trappe, M., Hipkins, V., Ye, T., Guida, S., Cronn, R., Cushman, S. A., & McEvoy, S. (2020). An Axiom SNP genotyping array for Douglas-fir. *BMC Genomics*, *21*(1), 9. https://doi.org/10.1186/s12864-019-6383-9

Hsu, F.-M., Yen, M.-R., Wang, C.-T., Lin, C.-Y., Wang, C.-J. R., & Chen, P.-Y. (2017). Optimized reduced representation bisulfite sequencing reveals tissue-specific mCHH islands in maize. *Epigenetics & Chromatin*, *10*(1), 42. https://doi.org/10.1186/s13072-017-0148-y

Hu, Y., Morota, G., Rosa, G. J. M., & Gianola, D. (2015). Prediction of Plant Height in Arabidopsis thaliana Using DNA Methylation Data. *Genetics*, *201*(2), 779–793. https://doi.org/10.1534/genetics.115.177204

Ibáñez, M. D., & Blázquez, M. A. (2019). Phytotoxic effects of commercial Eucalyptus citriodora, Lavandula angustifolia, and Pinus sylvestris essential oils on weeds, crops, and invasive species. *Molecules*, *24*(15), 2847. https://www.mdpi.com/1420-3049/24/15/2847

Kastally, C., Niskanen, A. K., Perry, A., Kujala, S. T., Avia, K., Cervantes, S., Haapanen, M., Kesälahti, R., Kumpula, T. A., Mattila, T. M., Ojeda, D. I., Tyrmi, J. S., Wachowiak, W., Cavers, S., Kärkkäinen, K., Savolainen, O., & Pyhäjärvi, T. (2022). Taming the massive genome of Scots pine with PiSy50k, a new genotyping array for conifer research. *The Plant Journal*, *109*(5), 1337–1350. https://doi.org/10.1111/tpj.15628

Kerfeld, C. A., & Scott, K. M. (2011). Using BLAST to Teach "E-value-tionary" Concepts. *PLoS Biology*, *9*(2), e1001014. https://doi.org/10.1371/journal.pbio.1001014

Klughammer, J., Datlinger, P., Printz, D., Sheffield, N. C., Farlik, M., Hadler, J., Fritsch, G., & Bock, C. (2015a). Differential DNA Methylation Analysis without a Reference Genome. *Cell Reports*, *13*(11), 2621–2633. https://doi.org/10.1016/j.celrep.2015.11.024

Klughammer, J., Datlinger, P., Printz, D., Sheffield, N. C., Farlik, M., Hadler, J., Fritsch, G., & Bock, C. (2015b). Differential DNA Methylation Analysis without a Reference Genome. *Cell Reports*, *13*(11), 2621–2633. https://doi.org/10.1016/j.celrep.2015.11.024

Komar, A. A. (Ed.). (2009). *Single Nucleotide Polymorphisms* (Vol. 578). Humana Press. https://doi.org/10.1007/978-1-60327-411-1

Krakau, U.-K., Liesebach, M., Aronen, T., Lelu-Walter, M.-A., & Schneck, V. (2013). Scots Pine (Pinus sylvestris L.). In L. E. Pâques (Ed.), *Forest Tree Breeding in Europe: Current State-of-the-Art and Perspectives* (pp. 267–323). Springer Netherlands. https://doi.org/10.1007/978-94-007-6146-9_6

Landi, M., & Guidi, L. (2023). Effects of abiotic stress on photosystem II proteins. *Photosynthetica*, *61*(SPECIAL ISSUE 2023/1), 148–156. https://doi.org/10.32615/ps.2022.043

Li, B., Zhou, W., Zhao, Y., Ju, Q., Yu, Z., Liang, Z., & Acharya, K. (2015). Using the SPEI to Assess Recent Climate Change in the Yarlung Zangbo River Basin, South Tibet. *Water*, *7*(10), Article 10. https://doi.org/10.3390/w7105474

Li, Q., Liu, C., Xin, Y., Shen, W., Zhao, W., Wang, F., Cao, Z., Bai, B., & Xin, P. (2024). Pinus Armandii Growth Gene Identification Using Genome-Wide Association Study Approaches. *Journal of Tropical Forest Science*, *36*(1), 26–39. https://www.jstor.org/stable/48758626

Lu, Y. (2016). Identification and Roles of Photosystem II Assembly, Stability, and Repair Factors in Arabidopsis. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.00168

Ma, X., Feng, F., Wei, H., Mei, H., Xu, K., Chen, S., Li, T., Liang, X., Liu, H., & Luo, L. (2016). Genome-Wide Association Study for Plant Height and Grain Yield in Rice under Contrasting Moisture Regimes. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.01801

Ma, Y., Liu, Y., Song, H., Sun, J., Lei, Y., & Wang, Y. (2015). A standardized precipitation evapotranspiration index reconstruction in the Taihe Mountains using tree-ring widths for the last 283 years. *PLoS One*, *10*(7), e0133605. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0133605

Malinowska, M., Nagy, I., Wagemaker, C. A. M., Ruud, A. K., Svane, S. F., Thorup-Kristensen, K., Jensen, C. S., Eriksen, B., Krusell, L., Jahoor, A., Jensen, J., Eriksen, L. B., & Asp, T. (2020). The cytosine methylation landscape of spring barley revealed by a new reduced representation bisulfite sequencing pipeline, WellMeth. *The Plant Genome*, *13*(3), e20049. https://doi.org/10.1002/tpg2.20049

Marek, S., Tomaszewski, D., Żytkowiak, R., Jasińska, A., Zadworny, M., Boratyńska, K., Dering, M., Danusevičius, D., Oleksyn, J., & Wyka, T. P. (2022). Stomatal density in

Pinus sylvestris as an indicator of temperature rather than CO_2: Evidence from a pan-European transect. *Plant, Cell & Environment*, *45*(1), 121–132. https://doi.org/10.1111/pce.14220

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., & Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, *33*(18), 5868–5877. https://doi.org/10.1093/nar/gki901

NIBIO_RAPPORT_2020_6_167.pdf. (n.d.). Retrieved September 29, 2023, from https://nibio.brage.unit.no/nibio-xmlui/bitstream/handle/11250/2720189/NIBIO_RAPPORT_2020_6_167.pdf?sequence=2&isAllowed=y

Oellrich, A., Walls, R. L., Cannon, E. K., Cannon, S. B., Cooper, L., Gardiner, J., Gkoutos, G. V., Harper, L., He, M., Hoehndorf, R., Jaiswal, P., Kalberer, S. R., Lloyd, J. P., Meinke, D., Menda, N., Moore, L., Nelson, R. T., Pujar, A., Lawrence, C. J., & Huala, E. (2015). An ontology approach to comparative phenomics in plants. *Plant Methods*, *11*(1), 10. https://doi.org/10.1186/s13007-015-0053-y

Ojeda, D. I., Mattila, T. M., Ruttink, T., Kujala, S. T., Kärkkäinen, K., Verta, J.-P., & Pyhäjärvi, T. (2019). Utilization of Tissue Ploidy Level Variation in de Novo Transcriptome Assembly of Pinus sylvestris. *G3 Genes|Genomes|Genetics*, *9*(10), 3409–3421. https://doi.org/10.1534/g3.119.400357

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, *21*(12), 2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x

Pasho, E., & Alla, A. Q. (2015). Climate impacts on radial growth and vegetation activity of two co-existing Mediterranean pine species. *Canadian Journal of Forest Research*, *45*(12), 1748–1756. https://doi.org/10.1139/cjfr-2015-0146

Paun, O., Verhoeven, K. J. F., & Richards, C. L. (2019). Opportunities and limitations of reduced representation bisulfite sequencing in plant ecological epigenomics. *New Phytologist*, *221*(2), 738–742. https://doi.org/10.1111/nph.15388

Peña-Gallardo, M., Vicente-Serrano, S. M., Camarero, J. J., Gazol, A., Sánchez-Salguero, R., Domínguez-Castro, F., El Kenawy, A., Beguería-Portugés, S., Gutiérrez, E., & De Luis, M. (2018). Drought sensitiveness on forest growth in peninsular Spain and the Balearic Islands. *Forests*, *9*(9), 524. https://www.mdpi.com/1999-4907/9/9/524

Perry, A., Wachowiak, W., Beaton, J., Iason, G., Cottrell, J., & Cavers, S. (2022a). Identifying and testing marker–trait associations for growth and phenology in three pine species: Implications for genomic prediction. *Evolutionary Applications*, *15*(2), 330–348. https://doi.org/10.1111/eva.13345

Perry, A., Wachowiak, W., Beaton, J., Iason, G., Cottrell, J., & Cavers, S. (2022b). Identifying and testing marker–trait associations for growth and phenology in three pine species: Implications for genomic prediction. *Evolutionary Applications*, *15*(2), 330–348. https://doi.org/10.1111/eva.13345

Racolta, A., Nodine, M. D., Davies, K., Lee, C., Rowe, S., Velazco, Y., Wellington, R., & Tax, F. E. (2018). A Common Pathway of Root Growth Control and Response to CLE Peptides Through Two Receptor Kinases in *Arabidopsis*. *Genetics*, *208*(2), 687–704. https://doi.org/10.1534/genetics.117.300148

Schmidt, M., Van Bel, M., Woloszynska, M., Slabbinck, B., Martens, C., De Block, M., Coppens, F., & Van Lijsebettens, M. (2017). Plant-RRBS, a bisulfite and next-generation sequencing-based methylome profiling method enriching for coverage of

cytosine positions. *BMC Plant Biology*, *17*(1), 115. https://doi.org/10.1186/s12870-017-1070-y

Shaikh, A. A., Chachar, S., Chachar, M., Ahmed, N., Guan, C., & Zhang, P. (2022). Recent Advances in DNA Methylation and Their Potential Breeding Applications in Plants. *Horticulturae*, *8*(7), Article 7. https://doi.org/10.3390/horticulturae8070562

Sheller, M., Tóth, E. G., Ciocîrlan, E., Mikhaylov, P., Kulakov, S., Kulakova, N., Melnichenko, N., Ibe, A., Sukhikh, T., & Curtu, A. L. (2023). Genetic Diversity and Population Structure of Scots Pine (Pinus sylvestris L.) in Middle Siberia. *Forests*, *14*(1), Article 1. https://doi.org/10.3390/f14010119

Solvin, T., & Sundheim Fløistad, I. (2023). *Statistics: Forest Seeds and Plants in the Nordic Region–Version 2023*. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1792548

Stevenson, D., & Zumajo-Cardona, C. (2018). From Plant Ontology to Gene Ontology and back. *Current Plant Biology*, *14*, 66–69. https://doi.org/10.1016/j.cpb.2018.09.009

Tabakova, M. A., Arzac, A., Martínez, E., & Kirdyanov, A. V. (2020). Climatic factors controlling Pinus sylvestris radial growth along a transect of increasing continentality in southern Siberia. *Dendrochronologia*, *62*, 125709. https://www.sciencedirect.com/science/article/pii/S1125786520300473

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, *20*(8), 467–484. https://doi.org/10.1038/s41576-019-0127-1

Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *The Plant Genome*, *14*(1), e20077. https://doi.org/10.1002/tpg2.20077

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 1–21. https://doi.org/10.1038/s43586-021-00056-9

Upadhyaya, H. D., Wang, Y.-H., Gowda, C. L. L., & Sharma, S. (2013). Association mapping of maturity and plant height using SNP markers with the sorghum mini core collection. *Theoretical and Applied Genetics*, *126*(8), 2003–2015. https://doi.org/10.1007/s00122-013-2113-x

Vacek, Z., Linda, R., Cukor, J., Vacek, S., Šim\uunek, V., Gallo, J., & Vančura, K. (2021). Scots pine (Pinus sylvestris L.), the suitable pioneer species for afforestation of reclamation sites? *Forest Ecology and Management*, *485*, 118951. https://www.sciencedirect.com/science/article/pii/S0378112721000402

Verissimo, A. F., & Daldal, F. (2014). Cytochrome *c* biogenesis System I: An intricate process catalyzed by a maturase supercomplex? *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1837*(7), 989–998. https://doi.org/10.1016/j.bbabio.2014.03.003

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010a). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010b). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Wang, Y., Zuo, L., Wei, T., Zhang, Y., Zhang, Y., Ming, R., Bachar, D., Xiao, W., Madiha, K., Chen, C., Fan, Q., Li, C., & Liu, J.-H. (2022). CHH methylation of genes associated

with fatty acid and jasmonate biosynthesis contributes to cold tolerance in autotetraploids of Poncirus trifoliata. *Journal of Integrative Plant Biology*, *64*(12), 2327–2343. https://doi.org/10.1111/jipb.13379

Waszak, N., Robertson, I., Puchałka, R., Przybylak, R., Pospieszyńska, A., & Koprowski, M. (2021). Investigating the climate-growth response of Scots Pine (Pinus sylvestris L.) in northern Poland. *Atmosphere*, *12*(12), 1690. https://www.mdpi.com/2073-4433/12/12/1690

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, *2*(3), 100141. https://doi.org/10.1016/j.xinn.2021.100141

Xiao, D., Zhou, K., Yang, X., Yang, Y., Ma, Y., & Wang, Y. (2021). Crosstalk of DNA Methylation Triggered by Pathogen in Poplars With Different Resistances. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.750089

Xiao, Q., Bai, X., Zhang, C., & He, Y. (2022). Advanced high-throughput plant phenotyping techniques for genome-wide association studies: A review. *Journal of Advanced Research*, *35*, 215–230. https://doi.org/10.1016/j.jare.2021.05.002

Yang, S., Gill, R. A., Zaman, Q. U., Ulhassan, Z., & Zhou, W. (2020). Insights on SNP types, detection methods and their utilization in Brassica species: Recent progress and future perspectives. *Journal of Biotechnology*, *324*, 11–20. https://www.sciencedirect.com/science/article/pii/S0168165620302546

Yuan, C., Xu, J., Chen, Q., Liu, Q., Hu, Y., Jin, Y., & Qin, C. (2021). C-terminal domain phosphatase-like 1 (CPL1) is involved in floral transition in Arabidopsis. *BMC Genomics*, *22*(1), 642. https://doi.org/10.1186/s12864-021-07966-8

Zhang, H., Lang, Z., & Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology*, *19*(8), 489–506. https://doi.org/10.1038/s41580-018-0016-z

Zhang, J., Yang, J., Zhang, L., Luo, J., Zhao, H., Zhang, J., & Wen, C. (2020). A new SNP genotyping technology Target SNP-seq and its application in genetic analysis of cucumber varieties. *Scientific Reports*, *10*(1), 5623. https://www.nature.com/articles/s41598-020-62518-6

Zhang, Y.-L., Jia, Q.-L., Li, D.-W., Wang, J.-E., Yin, Y.-X., & Gong, Z.-H. (2013). Characteristic of the Pepper CaRGA2 Gene in Defense Responses against Phytophthora capsici Leonian. *International Journal of Molecular Sciences*, *14*(5), Article 5. https://doi.org/10.3390/ijms14058985

Zhou, T., Yao, J., & Liu, Z. (2017). Gene Ontology, Enrichment Analysis, and Pathway Analysis. In Z. (John) Liu (Ed.), *Bioinformatics in Aquaculture* (1st ed., pp. 150–168). Wiley. https://doi.org/10.1002/9781118782392.ch10

Zhou, Y., Lei, Z., Zhou, F., Han, Y., Yu, D., & Zhang, Y. (2019). Impact of climate factors on height growth of Pinus sylvestris var. Mongolica. *PLoS One*, *14*(3), e0213509. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0213509&sid=QNQNb9

Appendix

Figure A1. EnrichGo command used in R-studio for gene ontology analysis.

```
library(clusterProfiler)
library(org.At.tair.db)
library(readxl)
library(svglite)
library(tidyverse)
library(RColorBrewer)
library(AnnotationDbi)
library(ggplot2)
library(enrichplot)

## setwd('/Users/prativapaudel/Documents/Gene ontology High SPEI LTHP(march-may)')

## GO_LTHP_RWC <- read.csv("GO_LTHP_RWC.csv", header = F,sep=",")
class(GO_LTHP_RWC)

##Convert DEG object from dataframe to character
'GO_LTHP_RWC' <- as.character(GO_LTHP_RWC[,1])

##Check if object is character
class(GO_LTHP_RWC)
head(GO_LTHP_RWC)

####Molecular function enrichment
EnrichGO_GO_LTHP_RWC <- enrichGO(GO_LTHP_RWC,
                        'org.At.tair.db',
                        pvalueCutoff=0.05,
                        qvalueCutoff =0.1,
                        pAdjustMethod="BH",
                        keyType = "SYMBOL",
                        ont="ALL")
as.data.frame(EnrichGO_GO_LTHP_RWC)

#####reduce redundancy of GO enriched terms
EnrichGO_GO_LTHP_RWC_redun <- clusterProfiler::simplify(EnrichGO_GO_LTHP_RWC,cutoff = 0.7, by =
"p.adjust",select_fun = min,measure = "Wang",semData = NULL)

#####Extract results
head(EnrichGO_GO_LTHP_RWC_redun@result)

##Since R identifies the GeneRatio, as character and not numeric, we need to format results to convert Ratio to fraction
EnrichGO_GO_LTHP_RWC_redun.df <- EnrichGO_GO_LTHP_RWC_redun@result %>%

  ##Separate ratios into 2 columns of data
  separate(BgRatio, into = c("size.term", "size.category"), sep = "/") %>%
  separate(GeneRatio, into = c("size.overlap.term", "size.overlap.category"),sep = "/") %>%

  ##Convert to numeric
mutate_at(vars("size.term", "size.category", "size.overlap.term","size.overlap.category"),as.numeric) %>%

###Calculate k/K
mutate(k.k=size.overlap.term/size.term)
View(EnrichGO_GO_LTHP_RWC_redun.df)

###Export enrichment results
write.csv(EnrichGO_GO_LTHP_RWC_redun.df, "EnrichGO_GO_LTHP_RWC_redun.df.csv")
barplot(EnrichGO_GO_HTLP_GWC_redun,showCategory = 10,font.size = 10)+ggtitle("EnrichGO for trees with Low SPEI")
dotplot(EnrichGO_GO_LTHP_RWC_redun,showCategory=10)+ggtitle("Enrich GO")
```

Figure A2. Admixture plots of Scots Pine generated by fastStructure software at different K value (2-7, 9 & 10).

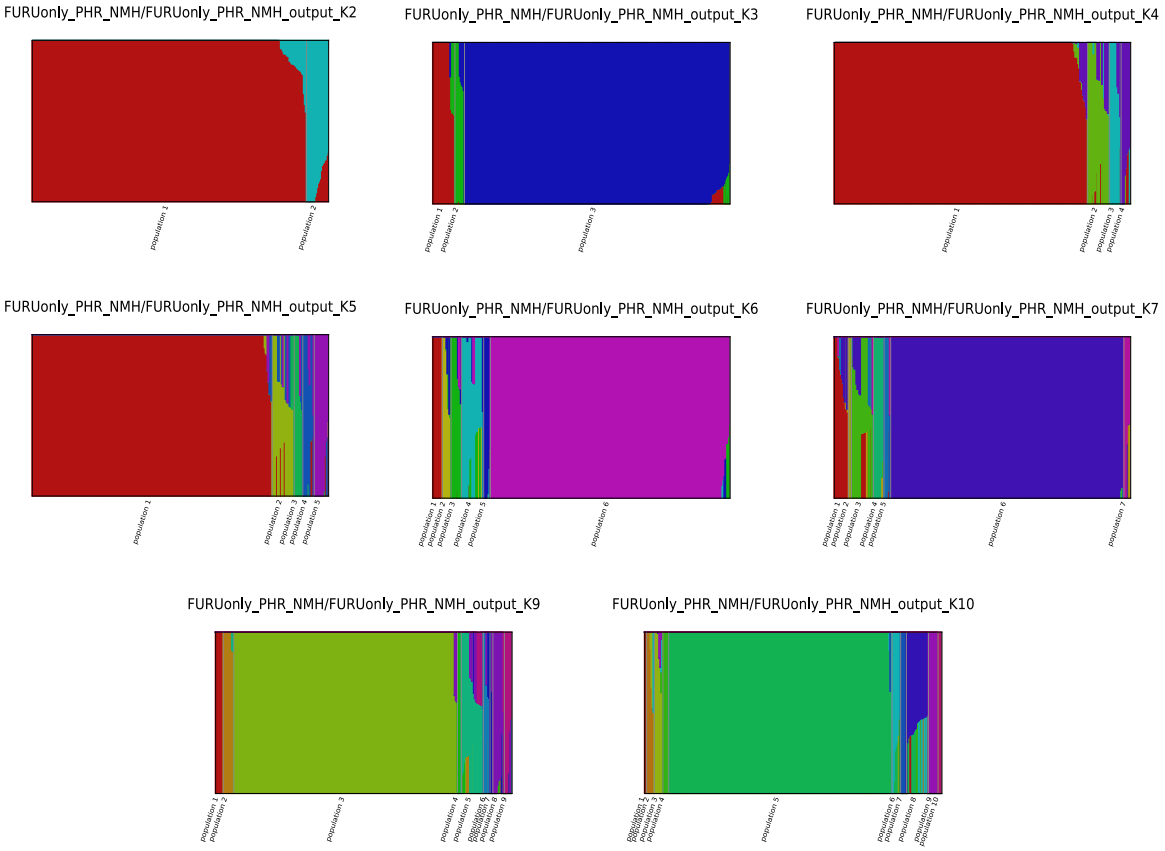
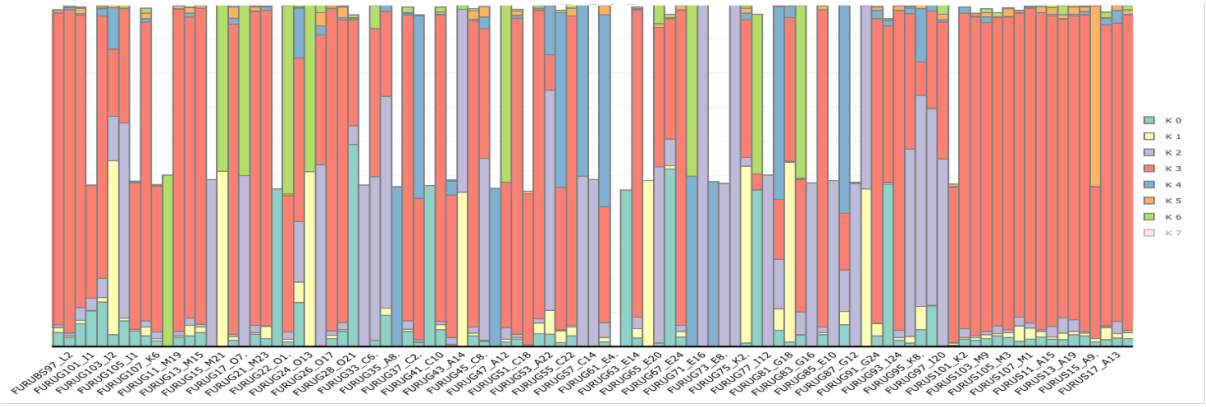


Figure A3. STRUCTURE analysis of 310 Scots Pine (Pinus sylvestris) individuals using 42,920 SNP loci reveals eight distinct clusters (K0 to K7), with each vertical line representing an individual and different colors indicating clusters.



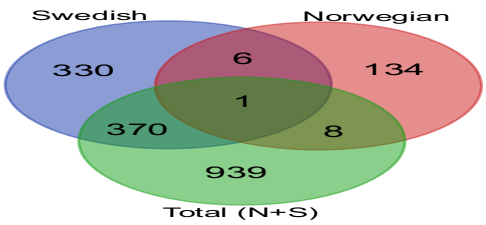


Figure A4. Venn diagram showing significantly associated SNPs of Norwegian, Swedish, and Total (N+S).

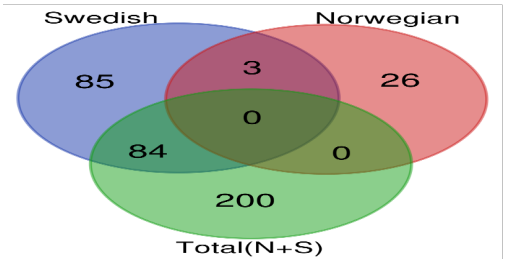


Figure A5. Venn diagram showing common genes identified in Norwegian, Swedish and Total (N+S).

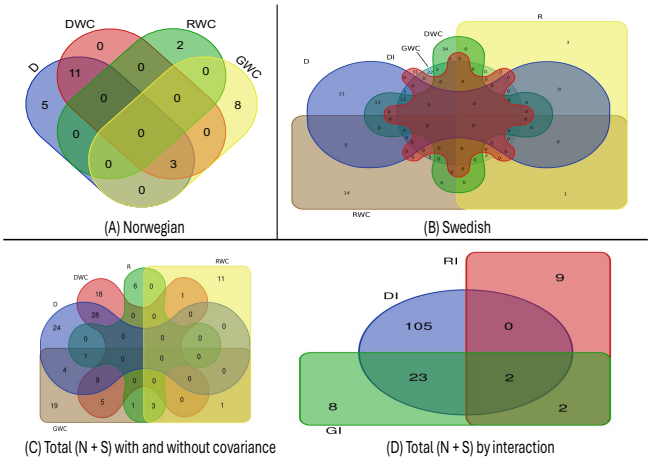


Figure A6. Venn diagram showing number of genes identified based on different Origin*. (A) represent genes identified based on the dominant, dominant interaction, and dominant, recessive, and genotype without covariance for Norwegian. (B) represent genes identified based on the dominant, dominant interaction, and dominant, recessive, and genotype without covariance for Swedish. (C) represent genes identified based on the dominant, recessive, and dominant, recessive, and genotype without covariance for Total(N+S). (D) represent genes identified based on the dominant, recessive, and genotype interaction for Total(N+S).*

Figure A7. Dot plots showing top 10 GO terms in the GO enrichment analysis of genes identified in Swedish trees. *(A) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant allele. (B) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant interaction. (C) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant allele without covariance.*

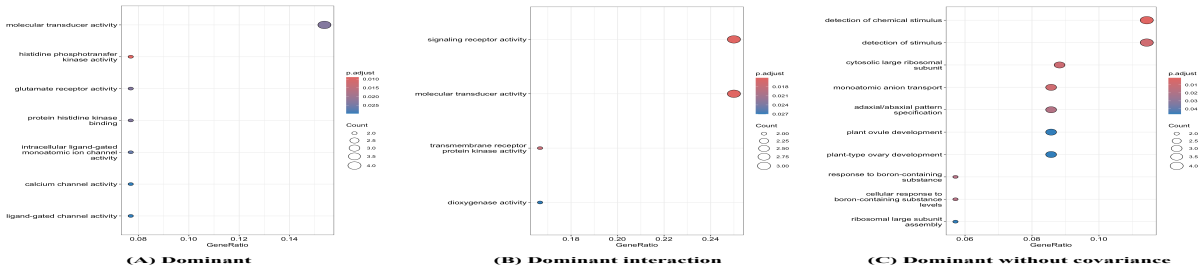


Figure A8. Dot plots showing top 10 GO terms in the GO enrichment analysis of genes identified in Swedish trees. *(A) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with recessive allele. (B) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with recessive without covariance. (C) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with genotype without covariance.*

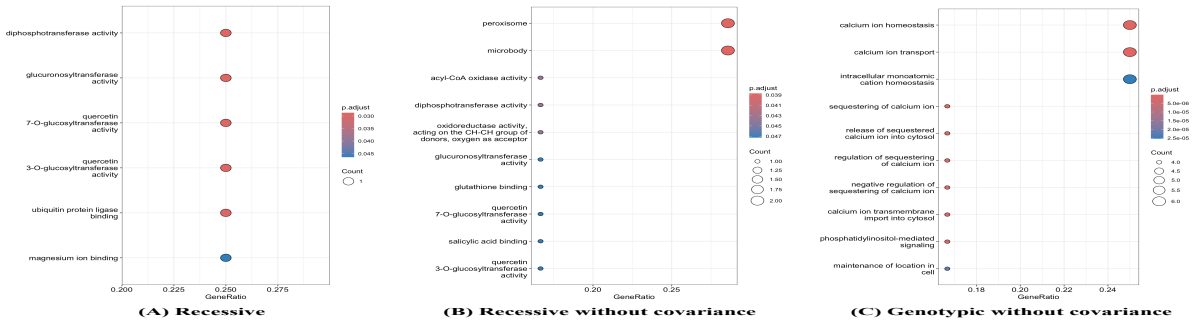


Figure A9. Dot plots showing top 10 GO terms in the GO enrichment analysis of genes identified in Total (N+S) trees. *(A) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant allele. (B) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant interaction. (C) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with dominant allele without covariance.*

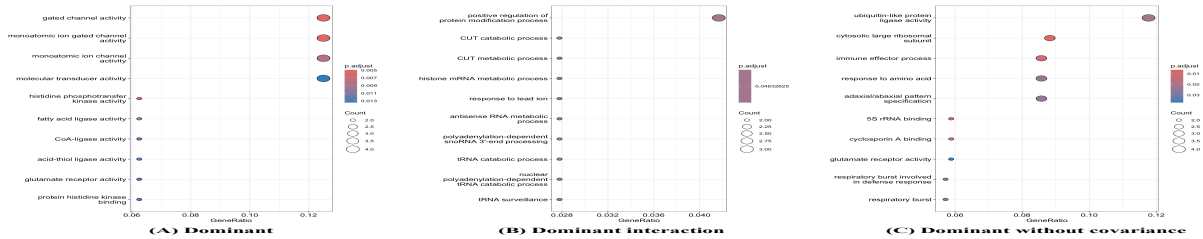


Figure A10. Dot plots showing top 10 GO terms in the GO enrichment analysis of genes identified in Total (N+S) trees. *(A) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with recessive allele. (B) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with recessive interaction. (C) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with recessive allele without covariance.*

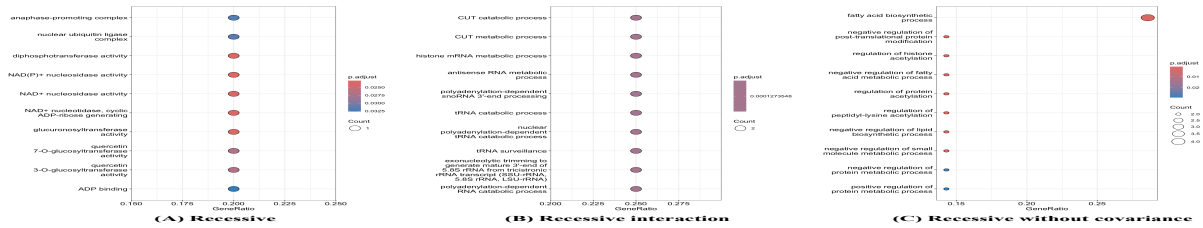
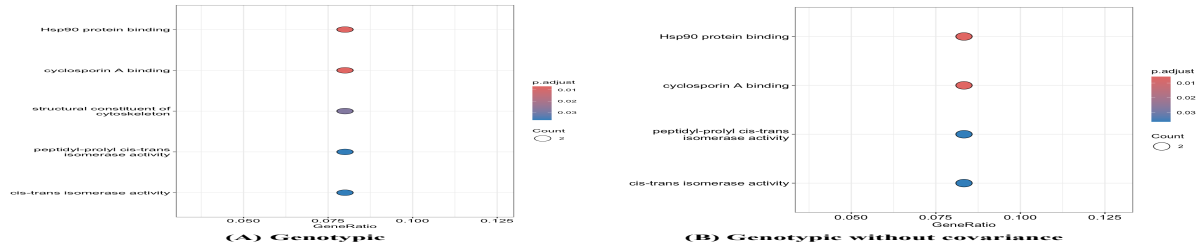


Figure A 11. Dot plots showing top 10 GO terms in the GO enrichment analysis of genes identified in Total (N+S) trees. *(A) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with genotype. (B) Dot plot of the top 10 GO terms in GO enrichment analysis for genes associated with genotype without covariance.*



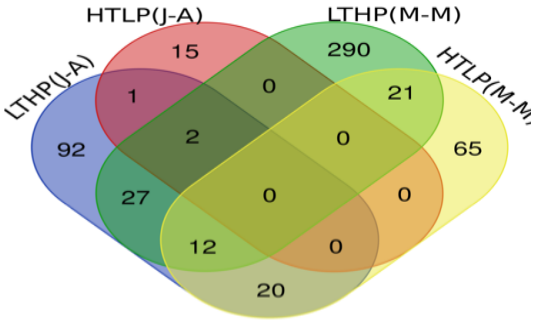


Figure A12. Venn diagram showing common genes between LTHP (March-May), HTLP (March-May), LTHP (June-August), and HTLP (June-August).

Figure A13. Principal component analysis (PCA 1, 2, 3, and 4) of DNA methylation profiles *based on CpG site of NvS for Read 1 and 2, along with clustering of 27 Pinus sylvestris individuals from 4 locations BM, BS, G, and S indicated by different color.*

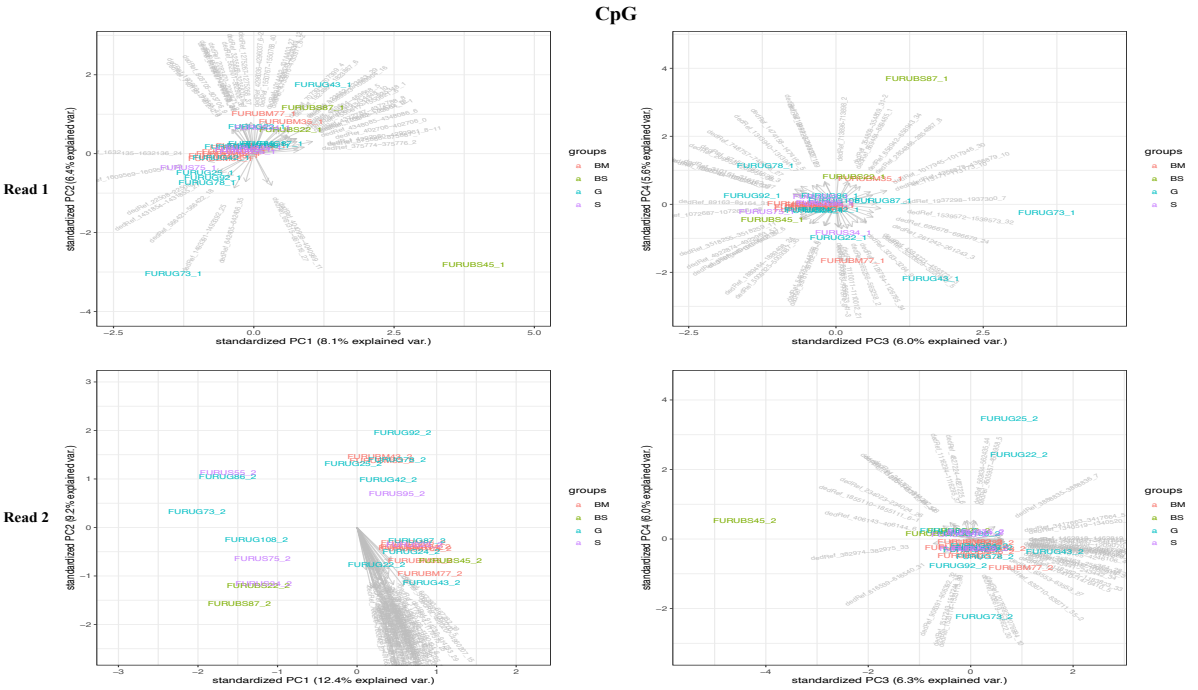


Figure A14. Principal component analysis (PCA 1, 2, 3, and 4) of DNA methylation profiles *based on CHG site of NvS for Read 1 and 2, along with clustering of 27 Pinus sylvestris individuals from 4 locations BM, BS, G, and S indicated by different color.*

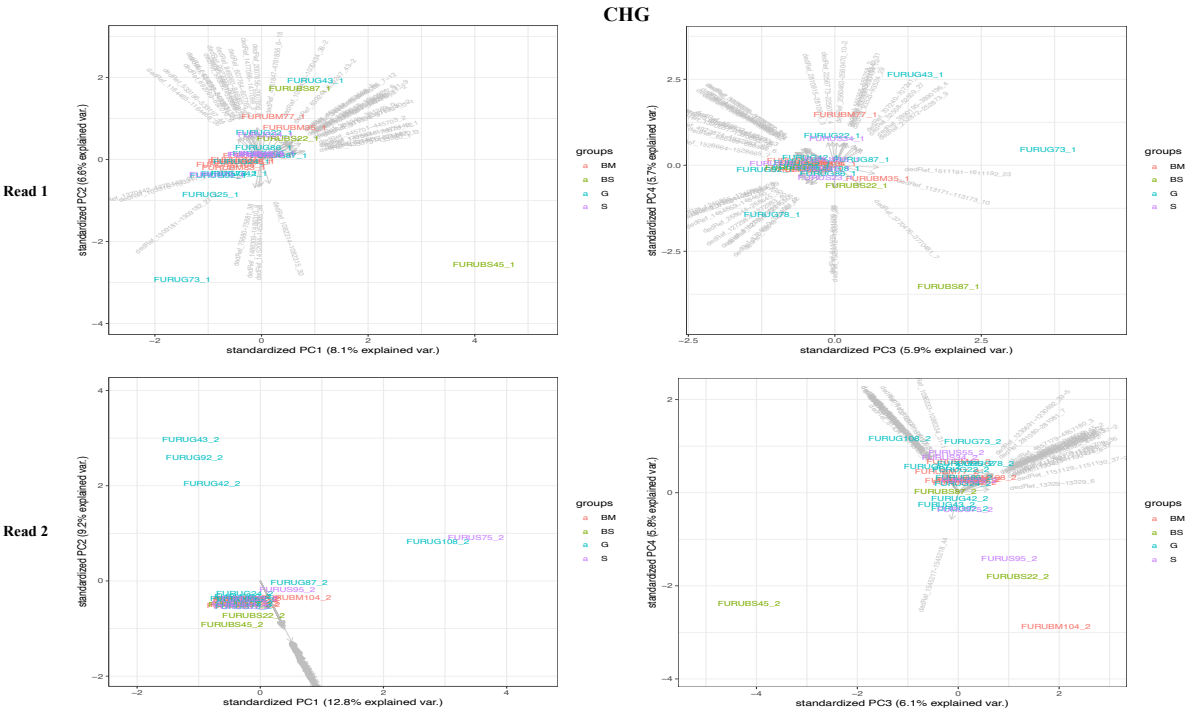


Figure A15. Principal component analysis (PCA 1, 2, 3, and 4) of DNA methylation profiles *based on CHH site of NvS for Read 1 and 2, along with clustering of 27 Pinus sylvestris individuals from 4 locations BM, BS, G, and S indicated by different color.*

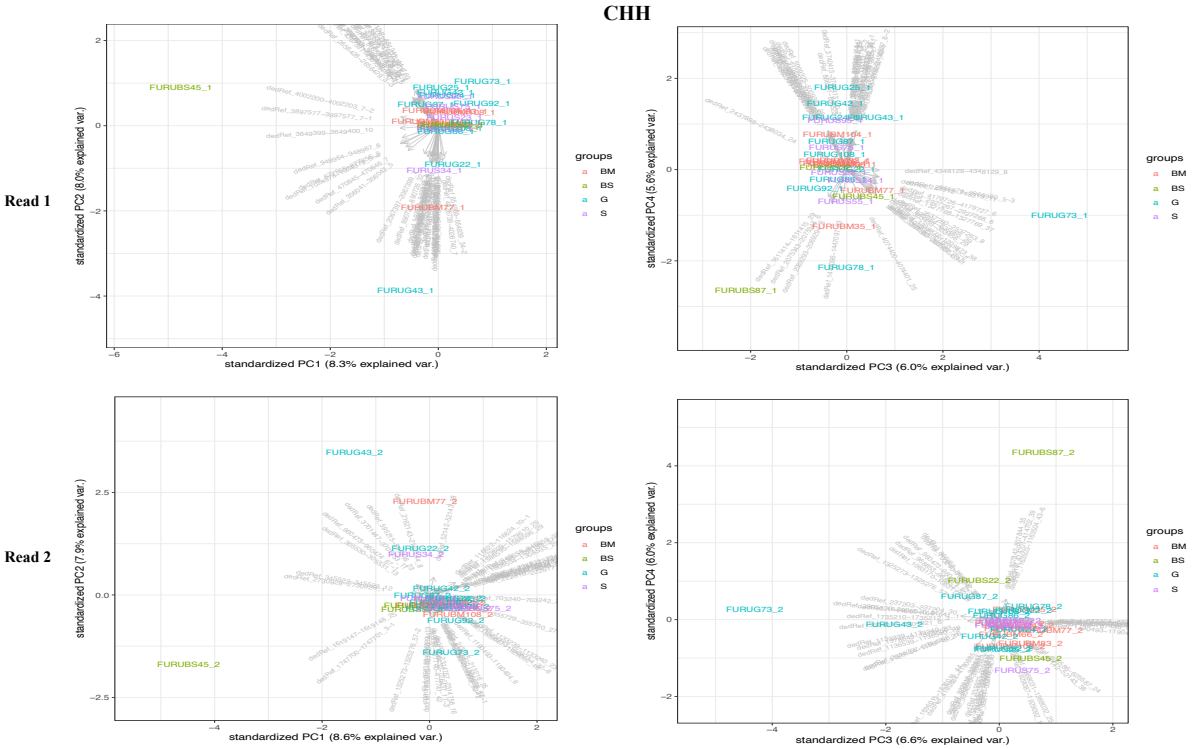


Figure A16. Principal component analysis (PCA 1, 2, 3, and 4) of DNA methylation profiles *based on CpG site of TvS for Read 1 and 2, along with clustering of 27 Pinus sylvestris individuals from 4 locations BM, BS, G, and S indicated by different color.*

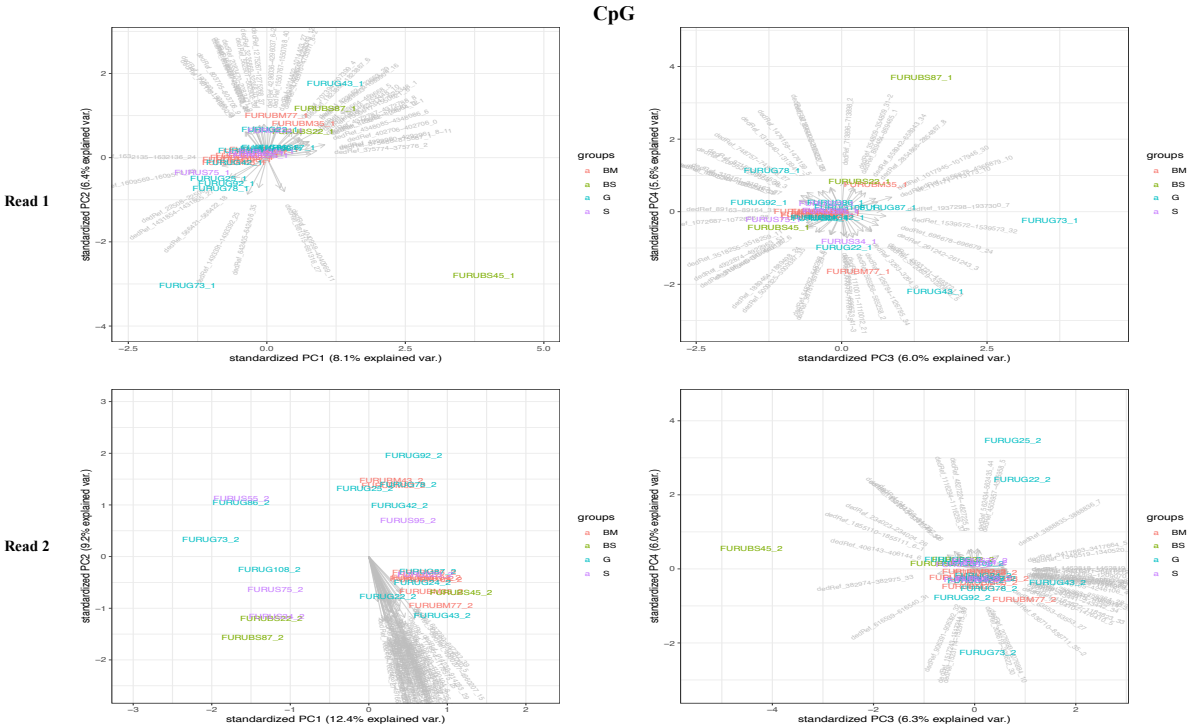


Figure A17. Principal component analysis (PCA 1, 2, 3, and 4) of DNA methylation profiles *based on CHG site of TvS for Read 1 and 2, along with clustering of 27 Pinus sylvestris individuals from 4 locations BM, BS, G, and S indicated by different color.*

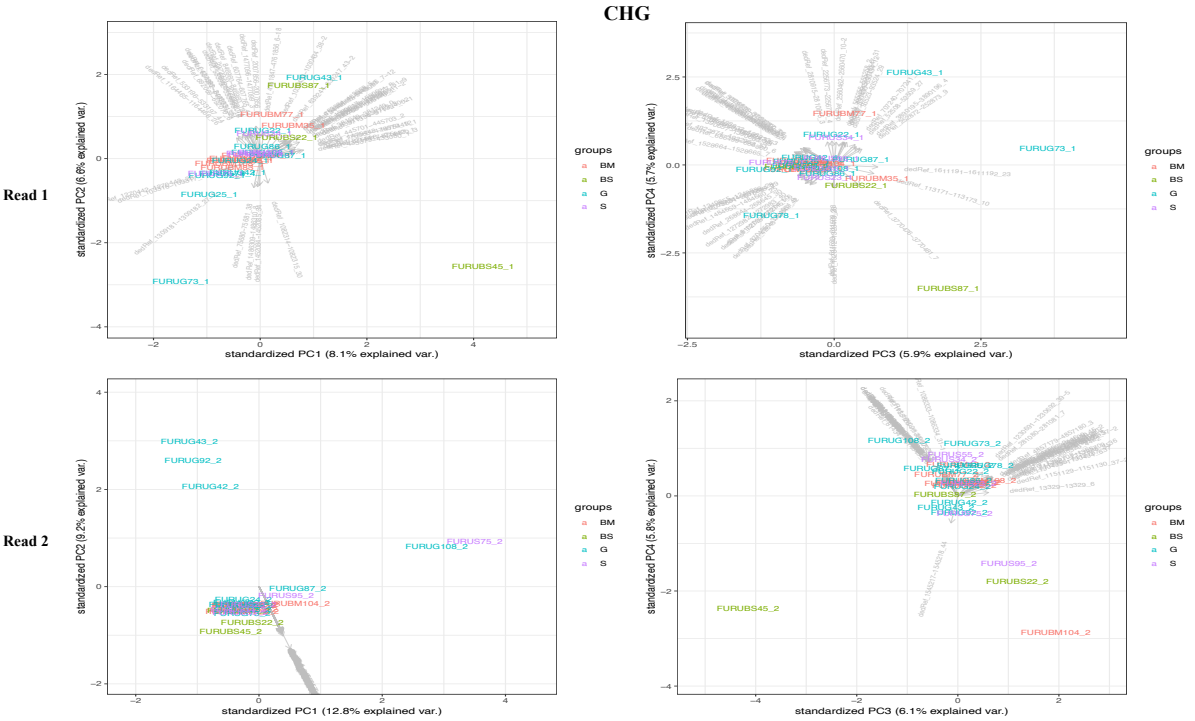


Figure A18. Principal component analysis (PCA 1, 2, 3, and 4) of DNA methylation profiles *based on CHH site of TvS for Read 1 and 2, along with clustering of 27 Pinus sylvestris individuals from 4 locations BM, BS, G, and S indicated by different color.*

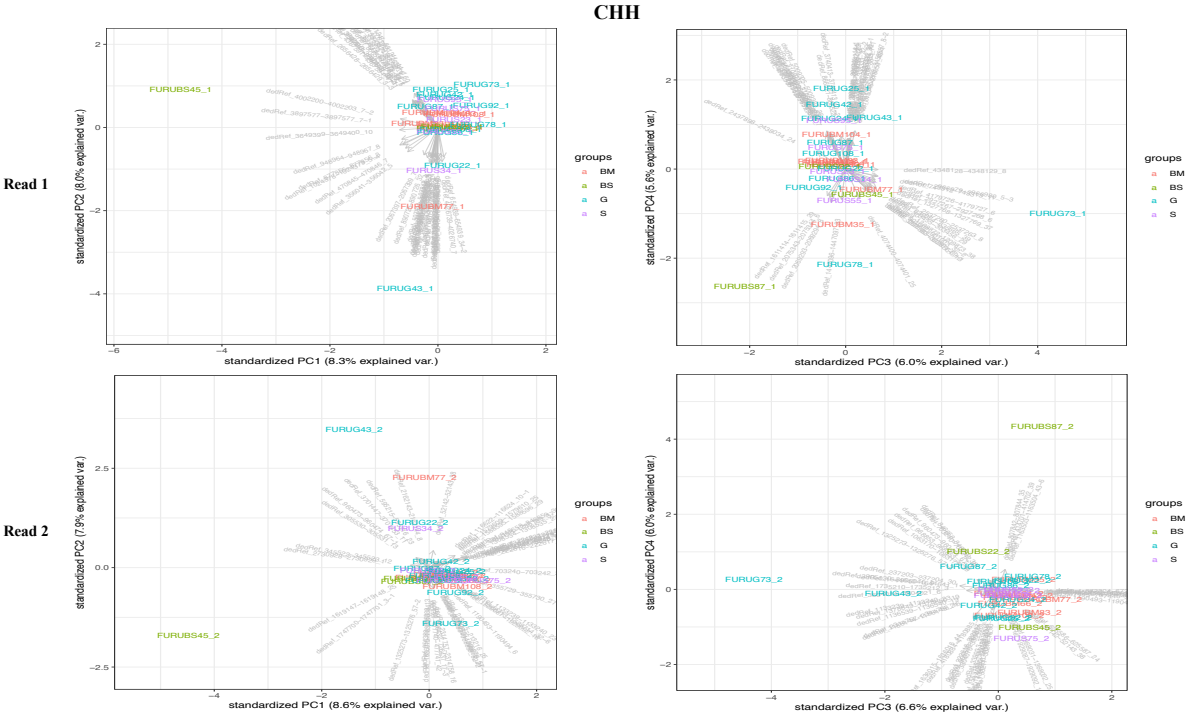
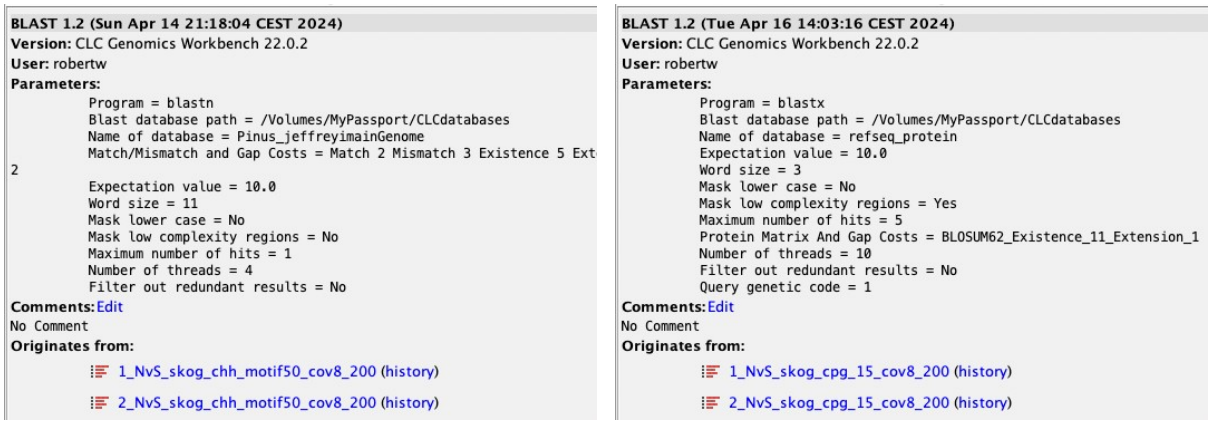


Figure A19. Parameters set during BLASTn and BLASTx search.



BLAST 1.2 (Sun Apr 14 21:18:04 CEST 2024)
Version: CLC Genomics Workbench 22.0.2
User: robertw
Parameters:
Program = blastn
Blast database path = /Volumes/MyPassport/CLCdatabases
Name of database = Pinus_jeffreyimainGenome
Match/Mismatch and Gap Costs = Match 2 Mismatch 3 Existence 5 Ext
2
Expectation value = 10.0
Word size = 11
Mask lower case = No
Mask low complexity regions = No
Maximum number of hits = 1
Number of threads = 4
Filter out redundant results = No
Comments: Edit
No Comment
Originates from:
1_NvS_skog_chh_motif50_cov8_200 (history)
2_NvS_skog_chh_motif50_cov8_200 (history)

(A)

BLAST 1.2 (Tue Apr 16 14:03:16 CEST 2024)
Version: CLC Genomics Workbench 22.0.2
User: robertw
Parameters:
Program = blastx
Blast database path = /Volumes/MyPassport/CLCdatabases
Name of database = refseq_protein
Expectation value = 10.0
Word size = 3
Mask lower case = No
Mask low complexity regions = Yes
Maximum number of hits = 5
Protein Matrix And Gap Costs = BLOSUM62_Existence_11_Extension_1
Number of threads = 10
Filter out redundant results = No
Query genetic code = 1
Comments: Edit
No Comment
Originates from:
1_NvS_skog_cpg_15_cov8_200 (history)
2_NvS_skog_cpg_15_cov8_200 (history)

(B)

Table S1. Total Samples (300) with annual growth measurement, total height (cm), and age.

Sample trees	Age	Height (cm) growth in individual yeras										Total Height (cm)
		2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	
FURUBM101_N13.CEL	7	20	25	9	13	7	7	6				87
FURUBM102_F5.CEL	8	44	36	18	25	25	35	23	14			220
FURUBM103_F13.CEL	8	44	37	28	27	24	28	14	12			214
FURUBM104_B15.CEL	8	51	37	31	42	37	33	22	26			279
FURUBM106_F17.CEL	7	34	21	11	13	17	22	5				123
FURUBM107_J23.CEL	9	30	26	18	22	17	29	19	9	3		173
FURUBM108_D17.CEL	9	57	58	37	40	35	37	26	14	10		314
FURUBM11_K24.CEL	8	47	37	26	25	18	24	17	8			202
FURUBM13_B3.CEL	8	45	39	34	38	28	28	17	5			234
FURUBM14_O2.CEL	8	45	38	29	30	25	29	19	5			220
FURUBM15_M16.CEL	8	30	33	27	29	26	16	9	3			173
FURUBM16_K12.CEL	9	22	16	9	13	15	21	8	3	2		109
FURUBM17_K14.CEL	7	17	16	9	19	9	5	4				79
FURUBM18_O12.CEL	9	18	13	10	16	11	17	13	8	5		111
FURUBM21_O14.CEL	8	36	29	27	39	26	22	14	7			200
FURUBM22_K22.CEL	8	38	18	15	26	21	17	8	3			146
FURUBM23_O18.CEL	9	24	19	14	14	13	21	10	4	3		122
FURUBM24_K20.CEL	8	53	44	29	36	37	35	22	12			268
FURUBM25_M8.CEL	8	47	34	23	28	25	25	13	13			208
FURUBM26_O16.CEL	8	42	30	22	24	22	32	19	4			195
FURUBM27_B5.CEL	8	36	26	18	23	18	25	5	4			155
FURUBM28_B7.CEL	10	45	36	28	24	21	22	18	21	19	4	238
FURUBM31_O4.CEL	9	39	33	20	20	14	21	12	20	2		181
FURUBM32_B1.CEL	6	10	12	6	10	9	3					50
FURUBM33_M4.CEL	7	11	13	8	8	7	8	2				57
FURUBM34_O6.CEL	9	49	41	34	34	26	31	18	13	24		270
FURUBM35_K18.CEL	7	8	5	5	5	3	5	3				34
FURUBM38_K16.CEL	7	22	20	13	16	10	10	7				98

FURUBM41_M18.CEL	8	40	30	20	26	18	22	14	14			184
FURUBM42_M20.CEL	7	14	15	9	10	11	13	3				75
FURUBM43_M14.CEL	7	44	45	29	32	23	22	25				220
FURUBM45_M2.CEL	8	25	23	16	18	19	19	6	4			130
FURUBM46_B9.CEL	8	47	34	18	31	20	33	20	13			216
FURUBM47_M24.CEL	7	25	19	15	16	20	17	5				117
FURUBM48_O8.CEL	8	44	36	27	32	24	28	16	3			210
FURUBM51_O10.CEL	8	37	28	14	16	13	18	12	6			144
FURUBM52_H23.CEL	8	41	31	20	28	29	30	14	5			198
FURUBM53_M12.CEL	8	48	39	29	24	30	36	28	19			253
FURUBM55_M6.CEL	8	53	51	41	37	36	24	14	13			269
FURUBM56_M22.CEL	6	7	9	5	10	8	2					41
FURUBM57_O20.CEL	8	35	28	18	29	16	15	14	2			157
FURUBM58_B11.CEL	7	20	15	12	13	13	14	5				92
FURUBM61_O24.CEL	8	24	23	15	20	14	18	6	4			124
FURUBM62_B17.CEL	9	35	35	33	40	32	34	19	9	7		244
FURUBM63_F9.CEL	9	44	30	20	24	27	19	18	8	4		194
FURUBM64_D11.CEL	9	52	36	37	43	36	22	24	19	2		271
FURUBM65_D19.CEL	8	50	38	29	35	37	34	20	16			259
FURUBM66_F7.CEL	7	9	10	8	11	6	4	2				50
FURUBM67_D3.CEL	7	16	9	11	13	7	7	5				68
FURUBM68_B21.CEL	8	41	38	39	41	18	13	14	16			220
FURUBM71_F15.CEL	10	48	45	39	40	28	30	29	19	20	12	310
FURUBM72_O22.CEL	10	36	31	27	28	23	25	15	9	10	4	208
FURUBM74_L24.CEL	8	29	18	13	20	18	24	20	13			155
FURUBM75_D9.CEL	7	11	15	10	11	6	5	6				64
FURUBM76_N24.CEL	7	36	33	24	19	12	16	18				158
FURUBM77_P24.CEL	7	46	44	38	48	23	13	8				220
FURUBM78_P18.CEL	7	22	23	16	13	18	15	8				115
FURUBM81_F1.CEL	9	44	37	32	33	23	11	15	10	4		209
FURUBM82_B13.CEL	9	56	48	35	39	33	42	35	20	8		316

FURUBM83_D5.CEL	9	59	42	35	40	33	28	21	11	4		273
FURUBM84_D1.CEL	9	38	31	22	22	20	23	11	7	5		179
FURUBM85_D15.CEL	8	43	25	21	31	24	19	7	6			176
FURUBM86_D21.CEL	8	30	31	12	10	8	17	10	10			128
FURUBM87_D13.CEL	8	43	36	23	30	18	26	21	15			212
FURUBM88_D7.CEL	9	48	41	20	23	19	28	15	4	6		204
FURUBM91_B23.CEL	9	50	43	31	34	27	23	18	21	8		255
FURUBM92_D23.CEL	8	41	33	22	24	17	22	12	9			180
FURUBM93_L9.CEL	8	28	27	7	16	10	18	9	8			123
FURUBM94_J19.CEL	8	50	39	30	27	21	27	17	11			222
FURUBM96_F11.CEL	8	49	37	23	34	30	40	29	21			263
FURUBM97_F19.CEL	8	52	39	26	24	17	29	21	17			225
FURUBM98_B19.CEL	9	38	30	24	25	17	17	10	9	8		178
FURUBS101_N22.CEL	10	27	22	18	18	11	14	15	17	16	9	167
FURUBS102_H20.CEL	10	33	29	19	19	13	18	15	15	21	7	189
FURUBS103_F22.CEL	10	14	12	10	13	9	14	10	11	16	9	118
FURUBS104_N20.CEL	10	20	12	7	12	12	22	17	18	21	9	150
FURUBS105_P22.CEL	8	18	11	8	10	8	7	4	4			70
FURUBS106_L22.CEL	9	38	25	17	18	13	13	14	10	4		152
FURUBS107_D24.CEL	9	50	45	29	34	31	39	24	13	8		273
FURUBS108_B20.CEL	9	20	13	9	12	8	16	14	5	4		101
FURUBS11_H17.CEL	9	42	29	15	23	27	30	10	5	3		184
FURUBS12_J13.CEL	9	28	23	13	20	14	17	14	5	3		137
FURUBS13_F23.CEL	8	37	25	24	33	27	23	19	11			199
FURUBS14_H7.CEL	9	32	27	20	22	15	22	12	5	5		160
FURUBS15_H9.CEL	9	51	33	29	32	26	32	20	15	14		252
FURUBS16_L21.CEL	10	60	47	36	35	33	32	23	20	17	13	316
FURUBS17_H19.CEL	9	42	26	17	26	17	17	13	5	7		170
FURUBS18_H15.CEL	9	29	17	7	14	16	26	22	7	5		143
FURUBS21_H1.CEL	9	17	18	10	14	11	14	8	7	6		105
FURUBS22_L7.CEL	8	11	8	6	6	4	13	12	5			65

FURUBS23_J7.CEL	10	42	32	26	29	23	33	30	21	13	8	257
FURUBS24_H11.CEL	10	52	34	35	39	32	38	34	25	21	9	319
FURUBS25_L1.CEL	10	49	45	38	43	33	42	32	26	17	6	331
FURUBS26_H13.CEL	9	31	25	18	25	19	18	15	22	19		192
FURUBS27_J21.CEL	8	13	9	6	16	14	11	9	6			84
FURUBS28_J5.CEL	10	34	23	15	16	11	13	17	18	21	11	179
FURUBS31_L17.CEL	10	29	23	15	21	14	15	13	13	22	15	180
FURUBS32_L5.CEL	10	36	33	22	34	21	17	15	25	22	13	238
FURUBS34_L19.CEL	10	45	40	29	29	22	21	18	24	17	9	254
FURUBS35_F20.CEL	10	36	30	21	18	11	16	14	16	17	14	193
FURUBS36_J20.CEL	10	34	36	26	32	20	24	18	21	22	13	246
FURUBS37_H21.CEL	10	41	34	20	18	10	22	21	20	17	6	209
FURUBS38_L13.CEL	10	22	18	10	8	7	18	16	13	6	2	120
FURUBS41_J17.CEL	10	30	28	13	12	6	13	15	14	16	10	157
FURUBS42_P11.CEL	9	36	29	22	22	13	23	20	10	2		177
FURUBS43_J3.CEL	9	29	23	10	18	16	22	20	10	6		154
FURUBS44_J11.CEL	7	11	14	9	11	10	13	6				74
FURUBS45_J1.CEL	8	15	14	5	6	8	6	6	5			65
FURUBS46_N3.CEL	9	38	31	17	18	17	20	19	24	17		201
FURUBS47_L23.CEL	8	42	28	18	15	12	19	17	21			172
FURUBS48_D22.CEL	9	37	30	16	19	18	25	12	10	14		181
FURUBS51_N23.CEL	10	48	39	30	32	19	23	17	23	26	18	275
FURUBS52_N17.CEL	10	53	46	31	31	21	27	21	26	29	12	297
FURUBS53_L3.CEL	9	22	27	15	20	18	25	18	13	10		168
FURUBS54_J9.CEL	10	50	48	37	37	16	17	20	23	16	6	270
FURUBS55_N7.CEL	9	36	32	20	23	11	10	20	15	9		176
FURUBS56_L15.CEL	10	41	40	28	34	25	32	27	16	17	11	271
FURUBS57_J22.CEL	10	43	41	29	30	18	17	16	28	27	12	261
FURUBS58_P7.CEL	10	35	29	20	20	11	15	14	20	14	6	184
FURUBS61_P1.CEL	10	22	17	12	11	9	10	10	10	10	6	117
FURUBS62_P15.CEL	9	44	39	21	12	17	21	17	5	5		181

FURUBS63_J15.CEL	9	48	42	21	23	14	20	17	20	18		223
FURUBS64_L11.CEL	9	17	18	12	18	10	7	10	11	12		115
FURUBS65_N21.CEL	10	52	51	39	41	30	22	19	23	23	11	311
FURUBS66_P3.CEL	9	28	25	14	16	9	12	13	17	8		142
FURUBS68_N9.CEL	10	42	42	23	22	12	19	21	21	21	10	233
FURUBS71_N1.CEL	9	39	33	18	21	17	22	19	22	13		204
FURUBS72_N5.CEL	9	24	20	19	21	18	20	12	18	16		168
FURUBS73_H3.CEL	9	30	19	17	19	13	16	17	16	16		163
FURUBS74_N19.CEL	8	11	13	9	13	8	10	10	9			83
FURUBS76_P13.CEL	10	48	35	16	16	13	20	22	24	12	6	212
FURUBS77_N15.CEL	9	38	25	15	20	11	16	8	14	25		172
FURUBS78_D20.CEL	10	36	28	19	20	13	12	10	11	11	10	170
FURUBS81_P17.CEL	9	37	29	19	25	22	29	28	18	22		229
FURUBS82_H5.CEL	10	45	42	22	22	20	28	22	13	12	10	236
FURUBS84_P5.CEL	9	46	41	29	31	21	16	26	26	11		247
FURUBS85_J24.CEL	10	56	47	31	31	20	28	28	25	17	15	298
FURUBS86_P9.CEL	10	28	22	11	10	10	12	11	9	11	7	131
FURUBS87_P20.CEL	9	14	12	7	11	10	15	7	2	4		82
FURUBS88_P23.CEL	9	25	17	13	13	8	14	13	6	5		114
FURUBS91_F24.CEL	10	51	41	30	26	14	15	12	20	22	13	244
FURUBS92_B22.CEL	10	34	30	15	21	11	17	12	10	13	9	172
FURUBS93_H24.CEL	9	29	24	8	5	6	6	9	8	7		102
FURUBS94_P19.CEL	10	52	44	43	44	34	37	30	21	18	15	338
FURUBS95_P21.CEL	9	56	41	28	32	33	38	24	17	29		298
FURUBS96_B24.CEL	10	33	29	24	32	19	16	13	15	12	6	199
FURUBS97_L20.CEL	10	30	20	15	17	14	17	15	25	19	11	183
FURUBS98_H22.CEL	8	21	17	12	17	8	13	12	6			106
FURUG101_I18.CEL	9	43	39	25	19	16	21	13	6	2		184
FURUG102_I14.CEL	8	21	22	14	14	12	15	4	4			106
FURUG103_I22.CEL	9	39	31	28	27	19	14	12	10	4		184
FURUG104_K4.CEL	10	43	41	35	34	27	23	17	16	3	5	244

FURUG105_I16.CEL	8	37	36	26	24	20	29	7	3			182
FURUG106_I8.CEL	8	45	41	38	40	39	18	13	5			239
FURUG107_K6.CEL	8	30	33	27	29	26	12	4	5			166
FURUG108_K10.CEL	9	41	42	31	35	30	35	20	10	4		248
FURUG11_M19.CEL	10	35	18	12	24	18	15	11	10	12	12	167
FURUG12_M13.CEL	10	51	44	28	33	23	24	20	26	20	12	281
FURUG13_M15.CEL	9	33	37	26	28	28	34	19	18	15		238
FURUG14_O3.CEL	7	37	36	21	20	16	13	10				153
FURUG15_M21.CEL	6	45	44	24	15	12	16					156
FURUG16_A2.CEL	5	17	13	12	12	5						59
FURUG17_O7.CEL	8	39	36	32	35	26	17	11	8			204
FURUG18_O15.CEL	8	41	42	30	33	28	17	11	4			206
FURUG21_O9.CEL	8	49	53	34	30	23	19	22	5			235
FURUG22_O1.CEL	7	40	38	35	45	33	23	15				229
FURUG23_O19.CEL	7	35	35	26	22	16	15	13				162
FURUG24_O13.CEL	9	53	46	36	42	30	38	16	10	3		274
FURUG25_O11.CEL	7	45	37	34	42	40	26	8				232
FURUG26_O17.CEL	6	13	26	23	19	4	4					89
FURUG27_O5.CEL	8	40	36	31	44	38	18	10	4			221
FURUG28_O21.CEL	9	37	34	23	26	17	26	13	20	10		206
FURUG41_C10.CEL	8	15	14	14	15	11	12	10	8			99
FURUG42_A16.CEL	7	12	10	6	8	4	9	2				51
FURUG43_A14.CEL	8	29	21	15	15	9	11	9	9			118
FURUG44_A20.CEL	7	20	22	28	10	13	5	9				107
FURUG45_C8.CEL	7	31	33	20	11	7	8	2				112
FURUG46_A18.CEL	8	41	26	16	19	11	12	10	8			143
FURUG47_A12.CEL	8	31	23	24	24	11	9	14	10			146
FURUG48_C4.CEL	9	32	26	25	26	14	11	10	14	9		167
FURUG51_C18.CEL	8	46	42	24	19	22	25	18	6			202
FURUG52_C20.CEL	8	38	38	30	23	19	22	12	3			185
FURUG53_A22.CEL	8	29	31	22	14	10	11	5	6			128

FURUG54_A24.CEL	9	44	32	23	18	7	22	11	7	4		168
FURUG55_C22.CEL	6	26	16	9	15	21	19					106
FURUG56_C16.CEL	7	46	48	35	39	26	21	10				225
FURUG57_C14.CEL	8	48	38	29	28	21	18	9	4			195
FURUG58_E12.CEL	7	42	35	29	30	25	19	3				183
FURUG61_E4.CEL	9	31	24	19	19	16	25	20	4	3		161
FURUG62_C24.CEL	8	36	31	24	33	18	28	19	10			199
FURUG63_E14.CEL	9	41	34	20	26	13	12	7	8	4		165
FURUG64_E2.CEL	8	49	32	30	33	19	22	20	6			211
FURUG65_E20.CEL	8	42	29	26	31	18	14	19	18			197
FURUG66_G4.CEL	8	28	27	22	23	20	12	9	6			147
FURUG67_E24.CEL	6	31	27	21	8	9	10					106
FURUG68_G8.CEL	8	40	35	25	25	19	16	6	3			169
FURUG71_E16.CEL	8	38	21	21	25	15	14	6	10			150
FURUG72_E18.CEL	8	37	33	30	35	22	18	20	5			200
FURUG73_E8.CEL	6	43	35	28	28	28	17					179
FURUG74_G2.CEL	8	33	32	20	22	9	12	9	17			154
FURUG75_K2.CEL	8	39	33	20	23	12	14	11	8			160
FURUG76_E22.CEL	9	40	39	31	37	11	14	16	16	4		208
FURUG77_I12.CEL	7	33	24	16	20	11	9	11				124
FURUG78_G10.CEL	8	31	18	15	17	15	20	6	6			128
FURUG81_G18.CEL	7	37	31	26	20	12	17	15				158
FURUG82_G6.CEL	7	22	20	12	17	13	8	4				96
FURUG83_G16.CEL	7	32	32	32	28	19	22	4				169
FURUG84_G20.CEL	6	13	14	7	8	7	4					53
FURUG85_E10.CEL	9	45	40	30	30	15	14	17	14	5		210
FURUG86_I4.CEL	9	37	30	15	20	9	12	6	11	3		143
FURUG87_G12.CEL	8	33	25	18	12	9	9	6	5			117
FURUG88_G14.CEL	8	49	52	32	38	22	23	13	4			233
FURUG91_G24.CEL	9	48	34	25	49	37	8	12	8	6		227
FURUG92_I2.CEL	8	54	56	48	50	33	13	9	15			278

FURUG93_I24.CEL	7	37	24	14	12	13	8	5				113
FURUG94_G22.CEL	8	56	59	45	45	22	22	10	5			264
FURUG95_K8.CEL	8	41	35	39	29	19	18	6	5			192
FURUG96_I6.CEL	7	49	49	37	23	16	19	9				202
FURUG97_I20.CEL	7	28	29	29	28	22	17	7				160
FURUG98_I10.CEL	6	26	21	13	16	14	15					105
FURUS101_K21.CEL	6	18	16	12	16	14	7					83
FURUS102_K11.CEL	6	12	12	8	3	4	11					50
FURUS103_M9.CEL	6	21	22	21	21	9	8					102
FURUS104_I21.CEL	6	39	24	17	18	8	10					116
FURUS105_M3.CEL	6	21	16	13	13	10	5					78
FURUS106_M7.CEL	7	34	29	18	14	10	11	6				122
FURUS107_M1.CEL	6	14	10	8	9	6	8					55
FURUS108_I19.CEL	7	26	18	14	18	9	8	4				97
FURUS11_A15.CEL	6	41	38	27	31	19	8					164
FURUS12_A3.CEL	7	41	34	33	36	20	6	4				174
FURUS13_A19.CEL	6	37	34	27	28	19	9					154
FURUS14_A23.CEL	5	12	11	7	9	14						53
FURUS15_A9.CEL	6	12	19	6	6	6	5					54
FURUS16_A17.CEL	7	19	16	17	14	9	8	4				87
FURUS17_A13.CEL	6	38	30	25	21	19	8					141
FURUS18_A7.CEL	6	18	18	16	11	9	17					89
FURUS21_A11.CEL	7	21	13	12	9	6	12	4				77
FURUS22_E3.CEL	7	21	29	34	35	21	18	6				164
FURUS23_A21.CEL	7	12	9	9	8	4	4	3				49
FURUS24_E11.CEL	7	11	8	7	11	10	8	3				58
FURUS25_C21.CEL	6	15	13	11	14	12	12					77
FURUS26_C1.CEL	6	19	111	14	17	16	5					182
FURUS27_I11.CEL	7	25	20	15	13	8	10	7				98
FURUS28_C3.CEL	8	49	47	40	38	31	28	14	6			253
FURUS31_C5.CEL	7	28	24	17	17	8	5	4				103

FURUS32_M17.CEL	6	44	42	24	27	11	15					163
FURUS33_A5.CEL	6	14	11	7	9	7	3					51
FURUS34_E9.CEL	7	47	51	43	39	22	15	6				223
FURUS35_A1.CEL	7	34	41	25	23	18	19	4				164
FURUS36_E13.CEL	6	32	26	17	19	9	15					118
FURUS37_C15.CEL	6	34	37	21	21	12	15					140
FURUS38_E1.CEL	6	27	21	19	25	15	8					115
FURUS41_C11.CEL	6	29	24	19	24	13	10					119
FURUS42_G3.CEL	7	22	20	7	10	8	10	5				82
FURUS43_G1.CEL	6	24	13	10	13	10	6					76
FURUS44_E7.CEL	6	30	21	16	28	8	7					110
FURUS45_C23.CEL	5	15	11	13	8	4						51
FURUS46_C17.CEL	8	49	45	40	43	23	20	12	8			240
FURUS47_C13.CEL	6	24	18	13	13	7	9					84
FURUS48_C9.CEL	7	38	30	23	26	23	17	12				169
FURUS51_E5.CEL	6	53	44	26	33	19	10					185
FURUS52_G5.CEL	7	48	43	27	36	18	12	4				188
FURUS53_E15.CEL	7	42	38	28	30	17	12	3				170
FURUS54_G7.CEL	4	11	8	7	10							36
FURUS55_K7.CEL	6	47	48	41	42	19	14					211
FURUS56_C7.CEL	6	50	42	22	27	20	12					173
FURUS57_G15.CEL	7	31	25	21	31	17	12	5				142
FURUS58_G13.CEL	7	27	20	16	13	10	12	4				102
FURUS61_E23.CEL	7	35	30	25	22	9	16	5				142
FURUS62_E19.CEL	6	18	8	9	18	9	8					70
FURUS63_E17.CEL	6	15	15	16	8	11	3					68
FURUS64_I13.CEL	6	21	13	10	11	8	3					66
FURUS65_E21.CEL	7	16	16	15	7	6	5	4				69
FURUS66_G11.CEL	7	32	36	26	30	21	19	5				169
FURUS67_I9.CEL	6	24	23	10	12	9	13					91
FURUS68_K1.CEL	7	15	12	7	8	11	3	3				59

FURUS71_G9.CEL	6	53	34	24	22	10	4					147
FURUS72_I7.CEL	7	53	47	34	37	21	9	5				206
FURUS73_I1.CEL	6	34	23	14	12	11	4					98
FURUS74_G21.CEL	7	38	19	21	22	21	13	5				139
FURUS75_G23.CEL	6	12	17	7	7	3	4					50
FURUS76_G19.CEL	7	44	49	36	35	27	5	5				201
FURUS77_I3.CEL	6	32	20	13	14	11	7					97
FURUS78_G17.CEL	7	45	39	32	35	26	19	3				199
FURUS81_I17.CEL	6	17	22	10	15	9	6					79
FURUS82_I5.CEL	6	41	50	39	38	15	15					198
FURUS83_C19.CEL	7	41	23	25	39	23	9	4				164
FURUS84_K3.CEL	5	37	41	32	27	18						155
FURUS85_K9.CEL	6	36	41	17	16	8	15					133
FURUS86_K5.CEL	6	21	20	9	12	6	6					74
FURUS87_I15.CEL	6	32	31	19	25	18	18					143
FURUS88_O23.CEL	6	33	26	17	17	13	5					111
FURUS91_M5.CEL	7	27	22	14	12	7	7	4				93
FURUS92_K19.CEL	6	17	12	13	8	6	4					60
FURUS93_K15.CEL	6	23	17	14	11	9	9					83
FURUS94_K17.CEL	7	22	16	13	18	10	9	5				93
FURUS95_K13.CEL	7	11	8	4	3	6	8	4				44
FURUS96_I23.CEL	8	48	33	10	24	10	14	5	5			149
FURUS97_M11.CEL	6	18	13	12	13	13	5					74
FURUS98_K23.CEL	6	41	23	18	28	19	15					144

Table S2. Significant SNPs with genes identified based on Origin of 300 Scots Pine (Pinus sylvestris) samples.

Origin	Dominant	Dominant Interaction	Dominant without covariance	Recessive	Recessive interaction	Recessive without covariance	Genotypic interaction	Genotypic without covariance
				Norwegian				
SNPs with genes identified	8	0	14	0	0	2	0	10
				Swedish				
SNPs with genes identified	38	18	57	5	0	16	0	42
				Total (Norwegian + Swedish)				
SNPs with genes identified	60	115	56	10	9	29	26	43

Table S3. Growth associated significant SNPs with genes identified based on seasonal temperature and precipitation for different time-period.

Time period with SPEI	Dominant without covariance	Recessive without covariance	Genotypic without covariance
	March-May		
HTLP (Low SPEI), 2019			
SNPs with identified genes	72	16	44
LTHP (High SPEI), 2017			
SNPs with identified genes	78	226	64
	Jun-August		
HTLP (Low SPEI), 2012			
SNPs with identified genes	6	10	3
LTHP (High SPEI), 2018			
SNPs with identified genes	108	23	73

* Significance determined at p-value ≤ 0.01

Table S4. Gene Ontology Analysis of genes corresponding to significant SNPs in association analysis for Norwegian samples.

Genes	ONTOLOGY	GO ID	Gene Ontology term	Size overlap term / Count	Size overlap category	p. adjust	Association
RPK2/SIA2	BP	GO:0009846	pollen germination	2	7	0.0405330538716311	Dominant
E2FE	BP	GO:0032876	negative regulation of DNA endoreduplication	1	7	0.0495629689636819	Dominant
E2FE	BP	GO:0008156	negative regulation of DNA replication	1	7	0.0495629689636819	Dominant
E2FE	BP	GO:2000104	negative regulation of DNA-templated DNA replication	1	7	0.0495629689636819	Dominant
RPK2/BAM1	BP	GO:0048443	stamen development	2	7	0.0289889408562439	Dominant without covariance
RPK2/BAM1	BP	GO:0048466	androecium development	2	7	0.0289889408562439	Dominant without covariance
ABCC5/NRAMP6	BP	GO:0030003	intracellular monoatomic cation homeostasis	2	7	0.0289889408562439	Dominant without covariance
MPK12	BP	GO:0071244	cellular response to carbon dioxide	1	7	0.0289889408562439	Dominant without covariance
ABCC5/NRAMP6	BP	GO:0006873	intracellular monoatomic ion homeostasis	2	7	0.0289889408562439	Dominant without covariance
MPK12	BP	GO:0080026	response to indolebutyric acid	1	7	0.0289889408562439	Dominant without covariance
RPK2/BAM1	BP	GO:0010073	meristem maintenance	2	7	0.0289889408562439	Dominant without covariance
MPK12/ABCC5	BP	GO:0010118	stomatal movement	2	7	0.0289889408562439	Dominant without covariance
NRAMP6	BP	GO:0015691	cadmium ion transport	1	7	0.0289889408562439	Dominant without covariance
ABCC5/NRAMP6	BP	GO:0055082	intracellular chemical homeostasis	2	7	0.0289889408562439	Dominant without covariance
RPK2/BAM1	BP	GO:0048438	floral whorl development	2	7	0.0289889408562439	Dominant without covariance
RPK2	BP	GO:0000578	embryonic axis specification	1	7	0.0353167039850309	Dominant without covariance
MPK12	BP	GO:0010037	response to carbon dioxide	1	7	0.0371477124200839	Dominant without covariance
ABCC5	BP	GO:0055075	potassium ion homeostasis	1	7	0.0402916472001392	Dominant without covariance
BAM 1,00	BP	GO:0005983	starch catabolic process	1	7	0.0424400527957071	Dominant without covariance
BAM 1,00	BP	GO:0009251	glucan catabolic process	1	7	0.0440032012578877	Dominant without covariance
BAM 1,00	BP	GO:0009934	regulation of meristem structural organization	1	7	0.0440032012578877	Dominant without covariance
NRAMP6	MF	GO:0015086	cadmium ion transmembrane transporter activity	1	7	0.042920502142744	Dominant without covariance
MPK12	MF	GO:0004707	MAP kinase activity	1	7	0.0458367797105691	Dominant without covariance
ABCC5	MF	GO:0008559	ABC-type xenobiotic transporter activity	1	7	0.0458367797105691	Dominant without covariance
RPK2	BP	GO:0010152	pollen maturation	1	7	0.0362841657812691	Dominant without covariance

RPK2	BP	GO:0009956	radial pattern formation	1	7	0.04531764088098	Dominant without covariance
RPK2/BAM1	BP	GO:0048653	anther development	2	7	0.0289889408562439	Dominant without covariance
SH3P2	BP	GO:0009920	cell plate formation involved in plant-type cell wall biogenesis	1	1	0.00926833319491722	Recessive without covariance
SH3P2	BP	GO:0007009	plasma membrane organization	1	1	0.00926833319491722	Recessive without covariance
SH3P2	BP	GO:0000919	cell plate assembly	1	1	0.00926833319491722	Recessive without covariance
SH3P2	BP	GO:0072583	clathrin-dependent endocytosis	1	1	0.00926833319491722	Recessive without covariance
SH3P2	BP	GO:0006898	receptor-mediated endocytosis	1	1	0.00926833319491722	Recessive without covariance
SH3P2	BP	GO:0051130	positive regulation of cellular component organization	1	1	0.0107845812521503	Recessive without covariance
SH3P2	BP	GO:0140014	mitotic nuclear division	1	1	0.0107845812521503	Recessive without covariance
SH3P2	BP	GO:0098657	import into cell	1	1	0.014512914209783	Recessive without covariance
SH3P2	BP	GO:0000280	nuclear division	1	1	0.016977707119722	Recessive without covariance
SH3P2	BP	GO:1903047	mitotic cell cycle process	1	1	0.0178390499127978	Recessive without covariance
SH3P2	CC	GO:0005776	autophagosome	1	1	0.0097126057808549	Recessive without covariance
SH3P2	CC	GO:0030136	clathrin-coated vesicle	1	1	0.0097126057808549	Recessive without covariance
SH3P2	CC	GO:0009504	cell plate	1	1	0.0097126057808549	Recessive without covariance
SH3P2	CC	GO:0090406	pollen tube	1	1	0.0097126057808549	Recessive without covariance
SH3P2	CC	GO:0030135	coated vesicle	1	1	0.0097126057808549	Recessive without covariance
SH3P2	CC	GO:0120025	plasma membrane bounded cell projection	1	1	0.0097126057808549	Recessive without covariance
SH3P2	MF	GO:0032266	phosphatidylinositol-3-phosphate binding	1	1	0.00762076982691817	Recessive without covariance
SH3P2	MF	GO:0043130	ubiquitin binding	1	1	0.00762076982691817	Recessive without covariance
SH3P2	MF	GO:0005546	phosphatidylinositol-4,5-bisphosphate binding	1	1	0.00762076982691817	Recessive without covariance
SH3P2	MF	GO:1902936	phosphatidylinositol bisphosphate binding	1	1	0.00762076982691817	Recessive without covariance
SH3P2	MF	GO:0032182	ubiquitin-like protein binding	1	1	0.00762076982691817	Recessive without covariance
SH3P2	MF	GO:1901981	phosphatidylinositol phosphate binding	1	1	0.00762076982691817	Recessive without covariance
SGR6/PYL4	CC	GO:0009705	plant-type vacuole membrane	2	6	0.0113387740799476	Genotypic without covariance
AVPL1	CC	GO:0000137	Golgi cis cisterna	1	6	0.0197030646467404	Genotypic without covariance
AVPL1	MF	GO:0009678	diphosphate hydrolysis-driven proton transmembrane transporter activity	1	6	0.0302215747690859	Genotypic without covariance
CPL1	MF	GO:0008420	RNA polymerase II CTD heptapeptide repeat phosphatase activity	1	6	0.0302215747690859	Genotypic without covariance
PYL4	MF	GO:0004864	protein phosphatase inhibitor activity	1	6	0.0302215747690859	Genotypic without covariance

PYL4	MF	GO:0010427	abscisic acid binding	1	6	0.0302215747690859	Genotypic without covariance
PYL4	MF	GO:0019212	phosphatase inhibitor activity	1	6	0.0302215747690859	Genotypic without covariance
AVPL1	MF	GO:0004427	inorganic diphosphate phosphatase activity	1	6	0.0302215747690859	Genotypic without covariance
PYL4	MF	GO:0019840	isoprenoid binding	1	6	0.0316445907022511	Genotypic without covariance
PYL4	MF	GO:0043178	alcohol binding	1	6	0.0316445907022511	Genotypic without covariance
CPL1	MF	GO:0140994	RNA polymerase II CTD heptapeptide repeat modifying activity	1	6	0.0316445907022511	Genotypic without covariance
PYL4	MF	GO:0042562	hormone binding	1	6	0.0323332562780103	Genotypic without covariance
BAM 1,00	MF	GO:0033612	receptor serine/threonine kinase binding	1	6	0.0484508054001727	Genotypic without covariance
PYL4	MF	GO:0044389	ubiquitin-like protein ligase binding	1	6	0.0484508054001727	Genotypic without covariance
CUL1	BP	GO:0010265	SCF complex assembly	1	7	0.0495629689636819	Dominant
CUL1	BP	GO:0010265	SCF complex assembly	1	7	0.0289889408562439	Dominant without covariance
ABCC5	BP	GO:0030007	intracellular potassium ion homeostasis	1	7	0.0495629689636819	Dominant
ABCC5	BP	GO:0030007	intracellular potassium ion homeostasis	1	7	0.0289889408562439	Dominant without covariance
RPK2	BP	GO:0009942	longitudinal axis specification	1	7	0.0495629689636819	Dominant
RPK2	BP	GO:0009942	longitudinal axis specification	1	7	0.0289889408562439	Dominant without covariance
ABCC5	BP	GO:1901527	abscisic acid-activated signalling pathway involved in stomatal movement	1	7	0.0495629689636819	Dominant
ABCC5	BP	GO:1901527	abscisic acid-activated signalling pathway involved in stomatal movement	1	7	0.0289889408562439	Dominant without covariance
ABCC5/BAM1	MF	GO:0038023	signalling receptor activity	2	7	0.042920502142744	Dominant without covariance
PYL4/BAM1	MF	GO:0038023	signalling receptor activity	2	6	0.0302215747690859	Genotypic without covariance
ABCC5/BAM1	MF	GO:0060089	molecular transducer activity	2	7	0.042920502142744	Dominant without covariance
PYL4/BAM1	MF	GO:0060089	molecular transducer activity	2	6	0.0302215747690859	Genotypic without covariance
BAM 1,00	MF	GO:0016161	beta-amylase activity	1	7	0.042920502142744	Dominant without covariance
BAM 1,00	MF	GO:0016161	beta-amylase activity	1	6	0.0302215747690859	Genotypic without covariance
BAM 1,00	MF	GO:0102229	amylopectin maltohydrolase activity	1	7	0.042920502142744	Dominant without covariance
BAM 1,00	MF	GO:0102229	amylopectin maltohydrolase activity	1	6	0.0302215747690859	Genotypic without covariance
BAM 1,00	MF	GO:0016160	amylase activity	1	7	0.042920502142744	Dominant without covariance
BAM 1,00	MF	GO:0016160	amylase activity	1	6	0.0302215747690859	Genotypic without covariance

Table S5. Gene Ontology Analysis of genes corresponding to significant SNPs in association analysis for Swedish samples.

Genes	ONTOLOGY	GO ID	Gene Ontology term	Size overlap term / Count	Size overlap category	p.adjust	Association
AHP1/ATHP2	MF	GO:0009927	histidine phosphotransfer kinase activity	2	26	0.00890309749591109	Dominant
AHP1/ATHP2/GLR3.1/GLR3.3	MF	GO:0060089	molecular transducer activity	4	26	0.0233743397026329	Dominant
AHP1/ATHP2	MF	GO:0043424	protein histidine kinase binding	2	26	0.0233743397026329	Dominant
ACR4/NIK3/THE1	MF	GO:0038023	signaling receptor activity	3	12	0.0150441459338126	Dominant interaction
ACR4/NIK3/THE1	MF	GO:0060089	molecular transducer activity	3	12	0.0150441459338126	Dominant interaction
ACR4/THE1	MF	GO:0019199	transmembrane receptor protein kinase activity	2	12	0.017904308621623	Dominant interaction
ANS/F6'H1	MF	GO:0051213	dioxygenase activity	2	12	0.0270324053419468	Dominant interaction
PCO4/PCO5/SIZ1/BOR1	BP	GO:0009593	detection of chemical stimulus	4	35	0.00119971527404993	Dominant without covariance
PCO4/PCO5/SIZ1/BOR1	BP	GO:0051606	detection of stimulus	4	35	0.00826607493163302	Dominant without covariance
BOR1/BOR4/BOR5	BP	GO:0006820	monoatomic anion transport	3	35	0.00826607493163302	Dominant without covariance
BOR1/BOR4	BP	GO:0010036	response to boron-containing substance	2	35	0.0185225219224622	Dominant without covariance
BOR1/BOR4	BP	GO:0080029	cellular response to boron-containing substance levels	2	35	0.0185225219224622	Dominant without covariance
RPL5A/RPL5B/LUG	BP	GO:0009955	adaxial/abaxial pattern specification	3	35	0.0185225219224622	Dominant without covariance
RPL5A/RPL5B	BP	GO:0000027	ribosomal large subunit assembly	2	35	0.0465138348029082	Dominant without covariance
SIZ1/MYB56/ATL72	BP	GO:0048481	plant ovule development	3	35	0.0465138348029082	Dominant without covariance
SIZ1/MYB56/ATL72	BP	GO:0035670	plant-type ovary development	3	35	0.0465138348029082	Dominant without covariance
RPL5/RPL5A/RPL5B	CC	GO:0022625	cytosolic large ribosomal subunit	3	34	0.00935445742877732	Dominant without covariance
BOR1/BOR4/BOR5	MF	GO:0005452	solute: inorganic anion antiporter activity	3	35	0.00021407630173300 1	Dominant without covariance
RPL5A/RPL5B	MF	GO:0008097	5S rRNA binding	2	35	0.014129143718872	Dominant without covariance
BOR1/BOR4	MF	GO:0015562	efflux transmembrane transporter activity	2	35	0.0340759839170525	Dominant without covariance
AHL23/AHL22	MF	GO:0000217	DNA secondary structure binding	2	35	0.0340759839170525	Dominant without covariance
RPL5/RPL5A/RPL5B	MF	GO:0003735	structural constituent of ribosome	3	35	0.0340759839170525	Dominant without covariance
PCO4/PCO5	MF	GO:0016702	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	2	35	0.0340759839170525	Dominant without covariance
BOR1/BOR4/BOR5	MF	GO:0015297	antiporter activity	3	35	0.0340759839170525	Dominant without covariance
GLR3.1/GLR3.3	MF	GO:0015276	ligand-gated monoatomic ion channel activity	2	35	0.0340759839170525	Dominant without covariance

CUL1	MF	GO:0031625	ubiquitin protein ligase binding	1	4	0.0294597491937438	Recessive
PRS4	MF	GO:0000287	magnesium ion binding	1	4	0.0464181471200038	Recessive
ACX4/PEX16	CC	GO:0005777	peroxisome	2	7	0.0387025089349737	Recessive without covariance
ACX4/PEX16	CC	GO:0042579	microbody	2	7	0.0387025089349737	Recessive without covariance
ACX4	MF	GO:0003997	acyl-CoA oxidase activity	1	6	0.0434687992811413	Recessive without covariance
ACX4	MF	GO:0016634	oxidoreductase activity, acting on the CH-CH group of donors, oxygen as acceptor	1	6	0.0434687992811413	Recessive without covariance
GSTU20	MF	GO:0043295	glutathione binding	1	6	0.0472606268664637	Recessive without covariance
ACX4	MF	GO:1901149	salicylic acid binding	1	6	0.0472606268664637	Recessive without covariance
PLC2/PLC4/PLC8/PLC7/GLR3.1/GLR3.3	BP	GO:0006873	intracellular monoatomic ion homeostasis	6	24	2.53588829902931e-05	Genotypic without covariance
PK1/ATPK2	BP	GO:0045727	positive regulation of translation	2	24	0.00844154082559766	Genotypic without covariance
PK1/ATPK2	BP	GO:0034250	positive regulation of amide metabolic process	2	24	0.00889246345089576	Genotypic without covariance
GLR3.1/GLR3.3	BP	GO:0071230	cellular response to amino acid stimulus	2	24	0.0134717231132706	Genotypic without covariance
GLR3.1/GLR3.3	BP	GO:0043200	response to amino acid	2	24	0.0338889726740765	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7/GLR3.1/GLR3.3	BP	GO:0006816	calcium ion transport	6	24	1.83073561150744e-07	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7/GLR3.1/GLR3.3	BP	GO:0030003	intracellular monoatomic cation homeostasis	6	24	2.53588829902931e-05	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	BP	GO:0032879	regulation of localization	4	24	0.00380476190200272	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	BP	GO:0048015	phosphatidylinositol-mediated signalling	4	24	1.65912296423998e-06	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	BP	GO:0051208	sequestering of calcium ion	4	24	6.70999411433369e-08	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	BP	GO:0051209	release of sequestered calcium ion into cytosol	4	24	6.70999411433369e-08	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	BP	GO:0051282	regulation of sequestering of calcium ion	4	24	6.70999411433369e-08	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	BP	GO:0051283	negative regulation of sequestering of calcium ion	4	24	6.70999411433369e-08	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	BP	GO:0051651	maintenance of location in cell	4	24	2.15437136593546e-05	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7/GLR3.1/GLR3.3	BP	GO:0055074	calcium ion homeostasis	6	24	1.20317353719638e-07	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	BP	GO:0070588	calcium ion transmembrane transport	4	24	2.53588829902931e-05	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	BP	GO:0097553	calcium ion transmembrane import into cytosol	4	24	6.70999411433369e-08	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7/GLR3.1/GLR3.3	BP	GO:0098771	inorganic ion homeostasis	6	24	7.96472031389336e-05	Genotypic without covariance
MUR3/GAUT11/HPAT3	CC	GO:0031985	Golgi cisterna	3	24	0.00982401343264397	Genotypic without covariance

MUR3/GAUT11/HPAT3	CC	GO:0005795	Golgi stack	3	24	0.00982401343264397	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	MF	GO:0004435	phosphatidylinositol phospholipase C activity	4	24	1.01263933048928e-07	Genotypic without covariance
PLC2/PLC4/PLC8/PLC7	MF	GO:0008081	phosphoric diester hydrolase activity	4	24	2.94827845993127e-05	Genotypic without covariance
AHL23/AHL22	MF	GO:0003680	minor groove of adenine-thymine-rich DNA binding	2	35	0.0340759839170525	Dominant without covariance
AHL23/AHL22	MF	GO:0003680	minor groove of adenine-thymine-rich DNA binding	2	24	0.00750108779028555	Genotypic without covariance
GLR3.1/GLR3.3	MF	GO:0005217	intracellular ligand-gated monoatomic ion channel activity	2	26	0.0262946079525596	Dominant
GLR3.1/GLR3.3	MF	GO:0005217	intracellular ligand-gated monoatomic ion channel activity	2	35	0.0340759839170525	Dominant without covariance
GLR3.1/GLR3.3	MF	GO:0005217	intracellular ligand-gated monoatomic ion channel activity	2	24	0.0112178294939552	Genotypic without covariance
GLR3.1/GLR3.3	MF	GO:0005262	calcium channel activity	2	26	0.0297182417834704	Dominant
GLR3.1/GLR3.3	MF	GO:0005262	calcium channel activity	2	35	0.0340759839170525	Dominant without covariance
GLR3.1/GLR3.3	MF	GO:0005262	calcium channel activity	2	24	0.0152314467742527	Genotypic without covariance
GLR3.1/GLR3.3	MF	GO:0008066	glutamate receptor activity	2	26	0.0233743397026329	Dominant
GLR3.1/GLR3.3	MF	GO:0008066	glutamate receptor activity	2	35	0.0340759839170525	Dominant without covariance
GLR3.1/GLR3.3	MF	GO:0008066	glutamate receptor activity	2	24	0.0102516294465573	Genotypic without covariance
GLR3.1/GLR3.3	MF	GO:0022834	ligand-gated channel activity	2	26	0.0297182417834704	Dominant
GLR3.1/GLR3.3	MF	GO:0022834	ligand-gated channel activity	2	35	0.0340759839170525	Dominant without covariance
GLR3.1/GLR3.3	MF	GO:0022834	ligand-gated channel activity	2	24	0.0152314467742527	Genotypic without covariance
UGT85A3	MF	GO:0015020	glucuronosyltransferase activity	1	4	0.0287242465807536	Recessive
UGT85A3	MF	GO:0015020	glucuronosyltransferase activity	1	6	0.0472606268664637	Recessive without covariance
PRS4	MF	GO:0016778	diphosphotransferase activity	1	4	0.0287242465807536	Recessive
PRS4	MF	GO:0016778	diphosphotransferase activity	1	6	0.0434687992811413	Recessive without covariance
UGT85A3	MF	GO:0080043	quercetin 3-O-glucosyltransferase activity	1	4	0.0287242465807536	Recessive
UGT85A3	MF	GO:0080043	quercetin 3-O-glucosyltransferase activity	1	6	0.0472606268664637	Recessive without covariance
UGT85A3	MF	GO:0080044	quercetin 7-O-glucosyltransferase activity	1	4	0.0287242465807536	Recessive
UGT85A3	MF	GO:0080044	quercetin 7-O-glucosyltransferase activity	1	6	0.0472606268664637	Recessive without covariance
PLP6/PLP7	MF	GO:0004620	phospholipase activity	2	35	0.0380817463165705	Dominant without covariance
PLC2/PLC4/PLC8/PLC7	MF	GO:0004620	phospholipase activity	4	24	2.94827845993127e-05	Genotypic without covariance

Table S6. Gene Ontology Analysis of genes corresponding to significant SNPs in association analysis for Total (Norwegian + Swedish) samples.

Genes	ONTOLOGY	GO ID	Gene Ontology term	Size overlap term / Count	Size overlap category	p.adjust	Association
SLAH2/CLC-F/GLR3.1/GLR3.3	MF	GO:0022836	gated channel activity	4	32	0.00490653211643329	Dominant
SLAH2/CLC-F/GLR3.1/GLR3.3	MF	GO:0022839	monoatomic ion gated channel activity	4	32	0.00490653211643329	Dominant
AHP1/AHP3	MF	GO:0009927	histidine phosphotransfer kinase activity	2	32	0.00509187352087977	Dominant
SLAH2/CLC-F/GLR3.1/GLR3.3	MF	GO:0005216	monoatomic ion channel activity	4	32	0.00801366126391258	Dominant
AAE7/AAE13	MF	GO:0015645	fatty acid ligase activity	2	32	0.0093672628670177	Dominant
AAE7/AAE13	MF	GO:0016405	CoA-ligase activity	2	32	0.0127814585354019	Dominant
AAE7/AAE13	MF	GO:0016878	acid-thiol ligase activity	2	32	0.0127814585354019	Dominant
AHP1/AHP3	MF	GO:0043424	protein histidine kinase binding	2	32	0.0127814585354019	Dominant
SLAH2/CLC-F	MF	GO:0008509	monoatomic anion transmembrane transporter activity	2	32	0.0154835393172079	Dominant
AHP1/AHP3/GLR3.1/GLR3.3	MF	GO:0060089	molecular transducer activity	4	32	0.0134164081892897	Dominant
FZR3/TPX2/BRCA1	BP	GO:0031401	positive regulation of protein modification process	3	72	0.0463262483147009	Dominant interaction
ACBP1/ACBP2	BP	GO:0010288	response to lead ion	2	72	0.0463262483147009	Dominant interaction
FZR3/MIEL1/TPX2/BRCA1	BP	GO:0051247	positive regulation of protein metabolic process	4	72	0.0463262483147009	Dominant interaction
BLH2/BLH4	BP	GO:0048363	mucilage pectin metabolic process	2	72	0.048051215885284	Dominant interaction
FZR3/TTM3/BRCA1	CC	GO:0000152	nuclear ubiquitin ligase complex	3	72	0.0316555861180195	Dominant interaction
IQD1/WRKY15/WRKY74/WRKY11/WRKY21/WRKY39	MF	GO:0005516	calmodulin binding	6	68	0.00840955874525533	Dominant interaction
ACBP1/ACBP2	MF	GO:0000062	fatty-acyl-CoA binding	2	68	0.0224260645976265	Dominant interaction
ACBP1/ACBP2	MF	GO:0120227	acyl-CoA binding	2	68	0.0224260645976265	Dominant interaction
ACBP1/ACBP2	MF	GO:1901567	fatty acid derivative binding	2	68	0.0224260645976265	Dominant interaction
SRO2/RCD1	MF	GO:0003950	NAD+ ADP-ribosyltransferase activity	2	68	0.0366210040349881	Dominant interaction
TTM3	CC	GO:0005680	anaphase-promoting complex	1	5	0.0317801164383721	Recessive
PRS4	MF	GO:0016778	diphosphotransferase activity	1	5	0.0228695052075852	Recessive
TAO1	MF	GO:0050135	NAD(P)+ nucleosidase activity	1	5	0.0228695052075852	Recessive
TAO1	MF	GO:0003953	NAD+ nucleosidase activity	1	5	0.0228695052075852	Recessive
TAO1	MF	GO:0061809	NAD+ nucleotidase, cyclic ADP-ribose generating	1	5	0.0228695052075852	Recessive
UGT85A3	MF	GO:0015020	glucuronosyltransferase activity	1	5	0.0228695052075852	Recessive
UGT85A3	MF	GO:0080044	quercetin 7-O-glucosyltransferase activity	1	5	0.0266346668418021	Recessive

UGT85A3	MF	GO:0080043	quercetin 3-O-glucosyltransferase activity	1	5	0.0266346668418021	Recessive
TAO1	MF	GO:0043531	ADP binding	1	5	0.0326644427185983	Recessive
RRP6L1/RRP6L2	BP	GO:0070918	regulatory ncRNA processing	2	8	0.00636181238819481	Recessive interaction
UGP1	BP	GO:0006011	UDP-glucose metabolic process	1	8	0.0127036868680328	Recessive interaction
ABCG40	BP	GO:0015692	lead ion transport	1	8	0.0130157144347934	Recessive interaction
VHA-d2	BP	GO:0007035	vacuolar acidification	1	8	0.0130409352586312	Recessive interaction
VHA-d2	BP	GO:0051452	intracellular pH reduction	1	8	0.0130409352586312	Recessive interaction
ABCG40/ABCG39	BP	GO:0000302	response to reactive oxygen species	2	8	0.0175829392665519	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0046700	heterocycle catabolic process	2	8	0.0186960039428179	Recessive interaction
ABCG40	BP	GO:0080168	abscisic acid transport	1	8	0.0186960039428179	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0042254	ribosome biogenesis	2	8	0.0186960039428179	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0044270	cellular nitrogen compound catabolic process	2	8	0.0186960039428179	Recessive interaction
ABCG40	BP	GO:0046864	isoprenoid transport	1	8	0.0216938761173796	Recessive interaction
ABCG40	BP	GO:0046865	terpenoid transport	1	8	0.0216938761173796	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0019439	aromatic compound catabolic process	2	8	0.0219797840984286	Recessive interaction
UGP1	BP	GO:0005977	glycogen metabolic process	1	8	0.0219797840984286	Recessive interaction
UGP1	BP	GO:0006112	energy reserve metabolic process	1	8	0.0219797840984286	Recessive interaction
UGP1/ABCG40	BP	GO:0031668	cellular response to extracellular stimulus	2	8	0.0219797840984286	Recessive interaction
UGP1/ABCG40	BP	GO:0071496	cellular response to external stimulus	2	8	0.0221505219634807	Recessive interaction
RRP6L1/RRP6L2	BP	GO:1901361	organic cyclic compound catabolic process	2	8	0.0221939396345991	Recessive interaction
RRP6L1	BP	GO:0006346	DNA methylation-dependent heterochromatin formation	1	8	0.0271296119788101	Recessive interaction
ABCG39	BP	GO:0042908	xenobiotic transport	1	8	0.0279700827218098	Recessive interaction
RRP6L1	BP	GO:0044030	regulation of DNA methylation	1	8	0.0279700827218098	Recessive interaction
ABCG40	BP	GO:0015850	organic hydroxy compound transport	1	8	0.0279700827218098	Recessive interaction
RRP6L1	BP	GO:0140718	facultative heterochromatin formation	1	8	0.0279700827218098	Recessive interaction
ABCG40	BP	GO:0010496	intercellular transport	1	8	0.0392131787200863	Recessive interaction
ABCG40	BP	GO:0015718	monocarboxylic acid transport	1	8	0.0452789577699014	Recessive interaction
ABCG40	BP	GO:0042631	cellular response to water deprivation	1	8	0.0468613949082055	Recessive interaction
ABCG40	BP	GO:0071462	cellular response to water stimulus	1	8	0.0468613949082055	Recessive interaction
ABCG40	BP	GO:0098739	import across plasma membrane	1	8	0.0484819432439376	Recessive interaction
UGP1	BP	GO:0052543	callose deposition in cell wall	1	8	0.0498976848829883	Recessive interaction
VHA-d2	CC	GO:0016471	vacuolar proton-transporting V-type ATPase complex	1	8	0.0170978205584357	Recessive interaction
ABCG34/ABCG40/ABCG39/VHA-d2	MF	GO:0042626	ATPase-coupled transmembrane transporter activity	4	8	0.00032506291277331 9	Recessive interaction

RRP6L1/RRP6L2	MF	GO:0000175	3'-5'-RNA exonuclease activity	2	8	0.0009630229622515 2	Recessive interaction
RRP6L1/RRP6L2	MF	GO:0004532	RNA exonuclease activity	2	8	0.0009630229622515 2	Recessive interaction
RRP6L1/RRP6L2	MF	GO:0016896	RNA exonuclease activity, producing 5'-phosphomonoesters	2	8	0.0009630229622515 2	Recessive interaction
RRP6L1/RRP6L2	MF	GO:0003727	single-stranded RNA binding	2	8	0.0025931906602261	Recessive interaction
UGP1	MF	GO:0003983	UTP:glucose-1-phosphate uridylyltransferase activity	1	8	0.0068794946144150 6	Recessive interaction
UGP1	MF	GO:0051748	UTP-monosaccharide-1-phosphate uridylyltransferase activity	1	8	0.0068794946144150 6	Recessive interaction
ABCG40	MF	GO:0015562	efflux transmembrane transporter activity	1	8	0.0241506316091161	Recessive interaction
VHA-d2	MF	GO:0042625	ATPase-coupled ion transmembrane transporter activity	1	8	0.0241506316091161	Recessive interaction
VHA-d2	MF	GO:0044769	ATPase activity, coupled to transmembrane movement of ions, rotational mechanism	1	8	0.0241506316091161	Recessive interaction
VHA-d2	MF	GO:0046961	proton-transporting ATPase activity, rotational mechanism	1	8	0.0241506316091161	Recessive interaction
RR21	MF	GO:0000156	phosphorelay response regulator activity	1	8	0.0301410456453292	Recessive interaction
RAR1/PUB23/GLR3.3	BP	GO:0002252	immune effector process	3	35	0.0112384272339697	Dominant without covariance
RAR1/PUB23	BP	GO:0002679	respiratory burst involved in defense response	2	35	0.0238889267898048	Dominant without covariance
RAR1/PUB23	BP	GO:0045730	respiratory burst	2	35	0.0238889267898048	Dominant without covariance
GLR3.1/GLR3.3/MKP1	BP	GO:0043200	response to amino acid	3	35	0.0238889267898048	Dominant without covariance
RPL5A/RPL5B/LUG	BP	GO:0009955	adaxial/abaxial pattern specification	3	35	0.0246862311949652	Dominant without covariance
RPL5/RPL5A/RPL5B	CC	GO:0022625	cytosolic large ribosomal subunit	3	34	0.0063689922919334 9	Dominant without covariance
RPL5A/RPL5B	MF	GO:0008097	5S rRNA binding	2	34	0.016131890942939	Dominant without covariance
SIZ1/ROC1/PUB23/PRT6	MF	GO:0061659	ubiquitin-like protein ligase activity	4	34	0.0215920925165603	Dominant without covariance
GLR3.1/GLR3.3	MF	GO:0005217	intracellular ligand-gated monoatomic ion channel activity	2	34	0.0371975731500519	Dominant without covariance
RPL5/RPL5A/RPL5B	MF	GO:0003735	structural constituent of ribosome	3	34	0.0371975731500519	Dominant without covariance
GLR3.1/GLR3.3	MF	GO:0015276	ligand-gated monoatomic ion channel activity	2	34	0.0371975731500519	Dominant without covariance
GLR3.1/GLR3.3	MF	GO:0022834	ligand-gated channel activity	2	34	0.0371975731500519	Dominant without covariance
BARD1/BRCA1	BP	GO:1901874	negative regulation of post-translational protein modification	2	14	0.0008308831028724 47	Recessive without covariance
BARD1/BRCA1	BP	GO:0035065	regulation of histone acetylation	2	14	0.0008308831028724 47	Recessive without covariance
BARD1/BRCA1	BP	GO:0045922	negative regulation of fatty acid metabolic process	2	14	0.0008308831028724 47	Recessive without covariance

BARD1/BRCA1	BP	GO:1901983	regulation of protein acetylation	2	14	0.00083088310287244 7	Recessive without covariance
BARD1/BRCA1	BP	GO:2000756	regulation of peptidyl-lysine acetylation	2	14	0.00083088310287244 7	Recessive without covariance
ACC1/BARD1/BRCA1/PEX 16	BP	GO:0006633	fatty acid biosynthetic process	4	14	0.00083088310287244 7	Recessive without covariance
BARD1/BRCA1	BP	GO:0051055	negative regulation of lipid biosynthetic process	2	14	0.0016598764291521 3	Recessive without covariance
BARD1/BRCA1	BP	GO:0062014	negative regulation of small molecule metabolic process	2	14	0.0030520633863023 4	Recessive without covariance
BARD1/BRCA1	BP	GO:0051248	negative regulation of protein metabolic process	2	14	0.028393270993726 6	Recessive without covariance
FIPS5/HEN2	BP	GO:0016441	post-transcriptional gene silencing	2	14	0.032306724561453 4	Recessive without covariance
ACC1/BARD1	BP	GO:0009933	meristem structural organization	2	14	0.034770914209355 4	Recessive without covariance
BARD1/BRCA1	BP	GO:0000724	double-strand break repair via homologous recombination	2	14	0.035244534571177 7	Recessive without covariance
MGL	BP	GO:0009068	aspartate family amino acid catabolic process	1	14	0.035244534571177 7	Recessive without covariance
BARD1/BRCA1	BP	GO:0000725	recombinational repair	2	14	0.035244534571177 7	Recessive without covariance
PAO5	BP	GO:0046208	spermine catabolic process	1	14	0.035244534571177 7	Recessive without covariance
BRCA1	BP	GO:0071479	cellular response to ionizing radiation	1	14	0.035244534571177 7	Recessive without covariance
ACC1	BP	GO:2001295	malonyl-CoA biosynthetic process	1	14	0.037452616342036 5	Recessive without covariance
BARD1/BRCA1	BP	GO:0010565	regulation of cellular ketone metabolic process	2	14	0.040228658861324 7	Recessive without covariance
MGL	BP	GO:0051289	protein homotetramerization	1	14	0.040518982916017 6	Recessive without covariance
MGL	BP	GO:0009092	homoserine metabolic process	1	14	0.040518982916017 6	Recessive without covariance
ACC1	BP	GO:2001293	malonyl-CoA metabolic process	1	14	0.040518982916017 6	Recessive without covariance
BARD1/BRCA1	BP	GO:0045944	positive regulation of transcription by RNA polymerase II	2	14	0.040518982916017 6	Recessive without covariance
ASHH2	BP	GO:0090548	response to nitrate starvation	1	14	0.040518982916017 6	Recessive without covariance
MIP2	BP	GO:0006624	vacuolar protein processing	1	14	0.042066059041834 7	Recessive without covariance
PAO5	BP	GO:0008215	spermine metabolic process	1	14	0.043553521927154 6	Recessive without covariance
TTM3/BARD1/BRCA1	CC	GO:0000151	ubiquitin ligase complex	3	14	0.011093165467088 1	Recessive without covariance
FIPS5/HEN2	CC	GO:0016607	nuclear speck	2	14	0.017921758850141 3	Recessive without covariance
TPR2/TPR3	MF	GO:0051879	Hsp90 protein binding	2	24	0.005670849111030 01	Genotypic without covariance
ABCG40	BP	GO:0010496	intercellular transport	1	8	0.039213178720086 3	Genotypic interaction

FAR5	BP	GO:0035336	long-chain fatty-acyl-CoA metabolic process	1	17	0.0440257273572053	Genotypic interaction
ABCG40	BP	GO:0015718	monocarboxylic acid transport	1	8	0.0452789577699014	Genotypic interaction
ACBP1	BP	GO:0055089	fatty acid homeostasis	1	17	0.0462185008805991	Genotypic interaction
ACBP1/MFT	BP	GO:0010029	regulation of seed germination	2	17	0.0462185008805991	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0071034	CUT catabolic process	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0071043	CUT metabolic process	2	72	0.0463262483147009	Genotypic interaction
FZR3/TPX2/BRCA1	BP	GO:0031401	positive regulation of protein modification process	3	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0008334	histone mRNA metabolic process	2	72	0.0463262483147009	Genotypic interaction
ACBP1/ACBP2	BP	GO:0010288	response to lead ion	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0042868	antisense RNA metabolic process	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0071051	polyadenylation-dependent snoRNA 3'-end processing	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0016078	tRNA catabolic process	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0071038	nuclear polyadenylation-dependent tRNA catabolic process	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0106354	tRNA surveillance	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0043633	polyadenylation-dependent RNA catabolic process	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0043634	polyadenylation-dependent ncRNA catabolic process	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0071029	nuclear ncRNA surveillance	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0071035	nuclear polyadenylation-dependent rRNA catabolic process	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0071046	nuclear polyadenylation-dependent ncRNA catabolic process	2	72	0.0463262483147009	Genotypic interaction
ABCG40	BP	GO:0042631	cellular response to water deprivation	1	8	0.0468613949082055	Genotypic interaction
ABCG40	BP	GO:0071462	cellular response to water stimulus	1	8	0.0468613949082055	Genotypic interaction
ACBP1/ACBP2	BP	GO:0010876	lipid localization	2	17	0.0471526176104414	Genotypic interaction
BLH2/BLH4	BP	GO:0048363	mucilage pectin metabolic process	2	72	0.048051215885284	Genotypic interaction
ABCG40	BP	GO:0098739	import across plasma membrane	1	8	0.0484819432439376	Genotypic interaction
UGP1	BP	GO:0052543	callose deposition in cell wall	1	8	0.0498976848829883	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0000467	exonucleolytic trimming to generate mature 3'-end of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	2	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	CC	GO:0000176	nuclear exosome (RNase complex)	2	8	9.49036543114154e-05	Genotypic interaction
RRP6L1/RRP6L2	CC	GO:0000178	exosome (RNase complex)	2	8	9.49036543114154e-05	Genotypic interaction
RRP6L1/RRP6L2	CC	GO:1905354	exoribonuclease complex	2	8	9.49036543114154e-05	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0070918	regulatory ncRNA processing	2	17	0.039257961124556	Genotypic interaction

VHA-d2	BP	GO:0007035	vacuolar acidification	1	17	0.0441028046785886	Genotypic interaction
VHA-d2	BP	GO:0051452	intracellular pH reduction	1	17	0.0441028046785886	Genotypic interaction
GLR3.1/GLR3.3	MF	GO:0008066	glutamate receptor activity	2	32	0.0127814585354019	Dominant
GLR3.1/GLR3.3	MF	GO:0008066	glutamate receptor activity	2	34	0.0371975731500519	Dominant without covariance
GLR3.1/GLR3.3	MF	GO:0005262	calcium channel activity	2	32	0.0201805318050439	Dominant
GLR3.1/GLR3.3	MF	GO:0005262	calcium channel activity	2	34	0.0371975731500519	Dominant without covariance
TTM3	CC	GO:0000152	nuclear ubiquitin ligase complex	1	5	0.0317801164383721	Recessive
TTM3/BARD1/BRCA1	CC	GO:0000152	nuclear ubiquitin ligase complex	3	14	9.63545526318309e-05	Recessive without covariance
ROC1/ROC2	MF	GO:0016018	cyclosporin A binding	2	34	0.016131890942939	Dominant without covariance
ROC1/ROC2	MF	GO:0016018	cyclosporin A binding	2	24	0.00567084911103001	Genotypic without covariance
ROC1/ROC2	MF	GO:0003755	peptidyl-prolyl cis-trans isomerase activity	2	34	0.0371975731500519	Dominant without covariance
ROC1/ROC2	MF	GO:0003755	peptidyl-prolyl cis-trans isomerase activity	2	24	0.0363259596532817	Genotypic without covariance
ROC1/ROC2	MF	GO:0016859	cis-trans isomerase activity	2	34	0.0371975731500519	Dominant without covariance
ROC1/ROC2	MF	GO:0016859	cis-trans isomerase activity	2	24	0.0363259596532817	Genotypic without covariance
BARD1/BRCA1	BP	GO:0051247	positive regulation of protein metabolic process	2	14	0.0291036022601861	Recessive without covariance
FZR3/MIEL1/TPX2/BRCA1	BP	GO:0051247	positive regulation of protein metabolic process	4	72	0.0463262483147009	Genotypic interaction
RRP6L1/RRP6L2	BP	GO:0071034	CUT catabolic process	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0071034	CUT catabolic process	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0071043	CUT metabolic process	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0071043	CUT metabolic process	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0106354	tRNA surveillance	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0106354	tRNA surveillance	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0008334	histone mRNA metabolic process	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0008334	histone mRNA metabolic process	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0016078	tRNA catabolic process	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0016078	tRNA catabolic process	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0071029	nuclear ncRNA surveillance	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0071029	nuclear ncRNA surveillance	2	8	0.00012735478470375 4	Recessive interaction

RRP6L1/RRP6L2	BP	GO:0071035	nuclear polyadenylation-dependent rRNA catabolic process	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0071035	nuclear polyadenylation-dependent rRNA catabolic process	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0071046	nuclear polyadenylation-dependent ncRNA catabolic process	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0071046	nuclear polyadenylation-dependent ncRNA catabolic process	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0043633	polyadenylation-dependent RNA catabolic process	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0043633	polyadenylation-dependent RNA catabolic process	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0043634	polyadenylation-dependent ncRNA catabolic process	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0043634	polyadenylation-dependent ncRNA catabolic process	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0000467	exonucleolytic trimming to generate mature 3'-end of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0000467	exonucleolytic trimming to generate mature 3'-end of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0042868	antisense RNA metabolic process	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0042868	antisense RNA metabolic process	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0071051	polyadenylation-dependent snoRNA 3'-end processing	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0071051	polyadenylation-dependent snoRNA 3'-end processing	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	BP	GO:0071038	nuclear polyadenylation-dependent tRNA catabolic process	2	72	0.0463262483147009	Dominant interaction
RRP6L1/RRP6L2	BP	GO:0071038	nuclear polyadenylation-dependent tRNA catabolic process	2	8	0.00012735478470375 4	Recessive interaction
RRP6L1/RRP6L2	CC	GO:0000176	nuclear exosome (RNase complex)	2	72	0.0316555861180195	Dominant interaction
RRP6L1/RRP6L2	CC	GO:0000176	nuclear exosome (RNase complex)	2	8	9.49036543114154e-05	Recessive interaction
RRP6L1/RRP6L2	CC	GO:0000178	exosome (RNase complex)	2	72	0.0316555861180195	Dominant interaction
RRP6L1/RRP6L2	CC	GO:0000178	exosome (RNase complex)	2	8	9.49036543114154e-05	Recessive interaction
VHA-e1/VHA-e2	CC	GO:0033179	proton-transporting V-type ATPase, V0 domain	2	72	0.0316555861180195	Dominant interaction
VHA-d2	CC	GO:0033179	proton-transporting V-type ATPase, V0 domain	1	8	0.0170978205584357	Recessive interaction
RRP6L1/RRP6L2	CC	GO:1905354	exoribonuclease complex	2	72	0.0316555861180195	Dominant interaction
RRP6L1/RRP6L2	CC	GO:1905354	exoribonuclease complex	2	8	9.49036543114154e-05	Recessive interaction

Table S7. Gene Ontology Analysis of genes corresponding to significant SNPs in association analysis of growth in term of SPEI value during the time-period March-May.

Genes	Ontology	GO ID	Gene Ontology term	Size overlap term / Count	Size overlap category	p.adjust	SPEI / Temperature and Precipitation	Association
PCAP1/MYB98/CLV2/LUG	BP	GO:0045595	regulation of cell differentiation	4	52	0.0271838422076058	High SPEI / LTHP	DWC
PCAP1/TFIIB/RAC6/RAC7/LUG	BP	GO:0022604	regulation of cell morphogenesis	5	52	0.0271838422076058	High SPEI / LTHP	DWC
PCAP1/CSI1/TFIIB/RAC6/RAC7/LUG	BP	GO:0022603	regulation of anatomical structure morphogenesis	6	52	0.0271838422076058	High SPEI / LTHP	DWC
ANP1/MKK1/MKK2	BP	GO:0000165	MAPK cascade	3	52	0.0271838422076058	High SPEI / LTHP	DWC
PCAP1/CSI1/RAC6/RAC7	BP	GO:0030865	cortical cytoskeleton organization	4	52	0.0379962671239935	High SPEI / LTHP	DWC
PCAP1/CSI1/RAC6/RAC7	BP	GO:0051493	regulation of cytoskeleton organization	4	52	0.0379962671239935	High SPEI / LTHP	DWC
ACBP2/PCAP1/SBP1/CSE/CSI1/LUG	BP	GO:0010038	response to metal ion	6	52	0.0379962671239935	High SPEI / LTHP	DWC
ATG18C/ATG18D	BP	GO:0034497	protein localization to phagophore assembly site	2	52	0.0379962671239935	High SPEI / LTHP	DWC
WOX8/RAC6/RAC7	BP	GO:0007163	establishment or maintenance of cell polarity	3	52	0.0427333644488191	High SPEI / LTHP	DWC
ATG18C/ATG18D/PCAP1	MF	GO:0080025	phosphatidylinositol-3,5-bisphosphate binding	3	52	0.00144477519907477	High SPEI / LTHP	DWC
MKK1/MKK2/MKK6	MF	GO:0004708	MAP kinase kinase activity	3	52	0.00405019348910528	High SPEI / LTHP	DWC
ATG18C/ATG18D	MF	GO:0032266	phosphatidylinositol-3-phosphate binding	2	52	0.0271476482980617	High SPEI / LTHP	DWC
SRS1/MYB56/ATL72/KAN2/SEU	BP	GO:0048440	carpel development	5	52	0.00885876135826332	Low SPEI / HTLP	DWC
SRS1/MYB56/ATL72/KAN2/SEU	BP	GO:0048467	gynoecium development	5	52	0.00885876135826332	Low SPEI / HTLP	DWC
UPS1/UPS5	BP	GO:0015855	pyrimidine nucleobase transport	2	52	0.00885876135826332	Low SPEI / HTLP	DWC
UPS1/UPS5	BP	GO:0015857	uracil transport	2	52	0.00885876135826332	Low SPEI / HTLP	DWC
MYB56/ATL72/KAN2/SEU	BP	GO:0048481	plant ovule development	4	52	0.00885876135826332	Low SPEI / HTLP	DWC
MYB56/ATL72/KAN2/SEU	BP	GO:0035670	plant-type ovary development	4	52	0.00885876135826332	Low SPEI / HTLP	DWC
ALA3/ALA9	BP	GO:0034204	lipid translocation	2	52	0.00900227028222802	Low SPEI / HTLP	DWC
ALA3/ALA9	BP	GO:0045332	phospholipid translocation	2	52	0.00900227028222802	Low SPEI / HTLP	DWC
MKP1/GLR3.1/GLR3.3/PUX10/SEU	BP	GO:0010243	response to organonitrogen compound	5	52	0.0107614664380553	Low SPEI / HTLP	DWC
ALA3/ALA9	BP	GO:0097035	regulation of membrane lipid distribution	2	52	0.0122598303417154	Low SPEI / HTLP	DWC
PUX10/DGAT2	BP	GO:0034389	lipid droplet organization	2	52	0.0153410074517151	Low SPEI / HTLP	DWC
CTF7/BARD1/BRCA1/RAD51B	BP	GO:0000724	double-strand break repair via homologous recombination	4	52	0.0153410074517151	Low SPEI / HTLP	DWC

CTF7/BARD1/BRCA1/RAD51B	BP	GO:0000725	recombinational repair	4	52	0.0153410074517151	Low SPEI / HTLP	DWC
ALA3/ALA9	BP	GO:1901703	protein localization involved in auxin polar transport	2	52	0.0160703285093218	Low SPEI / HTLP	DWC
ATL73/BARD1/BRCA1/PRT6	BP	GO:0019216	regulation of lipid metabolic process	4	52	0.0160703285093218	Low SPEI / HTLP	DWC
RPL5A/RPL5B/KAN2	BP	GO:0009955	adaxial/abaxial pattern specification	3	52	0.0167325751185803	Low SPEI / HTLP	DWC
MAKR1/MYB56/ROC1/RAV1	BP	GO:0009741	response to brassinosteroid	4	52	0.0209183112790456	Low SPEI / HTLP	DWC
RPL5A/RPL5B	BP	GO:0000027	ribosomal large subunit assembly	2	52	0.0278956008241408	Low SPEI / HTLP	DWC
GLR3.1/GLR3.3	BP	GO:0071230	cellular response to amino acid stimulus	2	52	0.0489655938183312	Low SPEI / HTLP	DWC
RPL5/RPL5A/RPL5B	CC	GO:0022625	cytosolic large ribosomal subunit	3	50	0.00862574638997222	Low SPEI / HTLP	DWC
VPS2.2/VPS2.3	CC	GO:0005771	multivesicular body	2	50	0.026709147548549	Low SPEI / HTLP	DWC
UPS1/UPS5	MF	GO:0005274	allantoin:proton symporter activity	2	52	0.00781729174264489	Low SPEI / HTLP	DWC
UPS1/UPS5	MF	GO:0015391	nucleobase:monoatomic cation symporter activity	2	52	0.00781729174264489	Low SPEI / HTLP	DWC
UPS1/UPS5	MF	GO:0015505	uracil:monoatomic cation symporter activity	2	52	0.00781729174264489	Low SPEI / HTLP	DWC
ALA3/ALA9	MF	GO:0140303	intramembrane lipid transporter activity	2	52	0.00781729174264489	Low SPEI / HTLP	DWC
ALA3/ALA9	MF	GO:0140326	ATPase-coupled intramembrane lipid transporter activity	2	52	0.00781729174264489	Low SPEI / HTLP	DWC
UPS1/UPS5	MF	GO:0005350	pyrimidine nucleobase transmembrane transporter activity	2	52	0.00781729174264489	Low SPEI / HTLP	DWC
UPS1/UPS5	MF	GO:0015210	uracil transmembrane transporter activity	2	52	0.00781729174264489	Low SPEI / HTLP	DWC
RPL5A/RPL5B	MF	GO:0008097	5S rRNA binding	2	52	0.0091407743136777	Low SPEI / HTLP	DWC
HMA5/ALA3/ALA9	MF	GO:0015662	P-type ion transporter activity	3	52	0.0091407743136777	Low SPEI / HTLP	DWC
HMA5/ALA3/ALA9	MF	GO:0140358	P-type transmembrane transporter activity	3	52	0.0091407743136777	Low SPEI / HTLP	DWC
GLR3.1/GLR3.3	MF	GO:0008066	glutamate receptor activity	2	52	0.0431354699257004	Low SPEI / HTLP	DWC
AHL10/RAD51B	MF	GO:0000217	DNA secondary structure binding	2	52	0.0468173483213023	Low SPEI / HTLP	DWC
GDPDL1/GDPDL4/GDPDL6/GDPDL7	MF	GO:0008889	glycerophosphodiester phosphodiesterase activity	4	144	0.00753696297817989	High SPEI /LTHP	RWC
ILL3/ILL5	MF	GO:0016810	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	2	13	0.025441326560056	Low SPEI / HTLP	RWC
SCRM	MF	GO:0001046	core promoter sequence-specific DNA binding	1	13	0.025441326560056	Low SPEI / HTLP	RWC
RBOHF	MF	GO:0016174	NAD(P)H oxidase H2O2-forming activity	1	13	0.0423890066658027	Low SPEI / HTLP	RWC
BARD1/BRCA1	BP	GO:0051055	negative regulation of lipid biosynthetic process	2	25	0.00535012667331132	Low SPEI / HTLP	GWC
PILS5/EMS1/CR4/RAV1	BP	GO:0090696	post-embryonic plant organ development	4	25	0.014710315687284	Low SPEI / HTLP	GWC

ATL73/BARD1/BRCA1	BP	GO:0010565	regulation of cellular ketone metabolic process	3	25	0.0154426559916529	Low SPEI / HTLP	GWC
BARD1/BRCA1	BP	GO:0006475	internal protein amino acid acetylation	2	25	0.0187011505402629	Low SPEI / HTLP	GWC
PCO4/PCO5	BP	GO:0009593	detection of chemical stimulus	2	25	0.0187011505402629	Low SPEI / HTLP	GWC
EMS1/KAN2	BP	GO:0001708	cell fate specification	2	25	0.0261759753552301	Low SPEI / HTLP	GWC
PILS5/EMS1/CR4	BP	GO:1905393	plant organ formation	3	25	0.0261759753552301	Low SPEI / HTLP	GWC
HSP18.2/ROC1/ROC2	BP	GO:0006457	protein folding	3	25	0.0283137047504845	Low SPEI / HTLP	GWC
ATL73/BARD1/BRCA1	BP	GO:0042180	cellular ketone metabolic process	3	25	0.0283137047504845	Low SPEI / HTLP	GWC
PCO4/PCO5/RAV1	BP	GO:0071456	cellular response to hypoxia	3	25	0.0314225709431548	Low SPEI / HTLP	GWC
PCO4/PCO5/RAV1	BP	GO:0036294	cellular response to decreased oxygen levels	3	25	0.0314225709431548	Low SPEI / HTLP	GWC
PCO4/PCO5/RAV1	BP	GO:0071453	cellular response to oxygen levels	3	25	0.0314225709431548	Low SPEI / HTLP	GWC
BARD1/BRCA1	BP	GO:0051248	negative regulation of protein metabolic process	2	25	0.0463224989219289	Low SPEI / HTLP	GWC
BRCA1	BP	GO:0071479	cellular response to ionizing radiation	1	25	0.0463224989219289	Low SPEI / HTLP	GWC
CR4	BP	GO:0090392	sepal giant cell differentiation	1	25	0.0463224989219289	Low SPEI / HTLP	GWC
PCO4/PCO5/RAV1	BP	GO:0036293	response to decreased oxygen levels	3	25	0.0463224989219289	Low SPEI / HTLP	GWC
HSP18.2/ROC1/ROC2	BP	GO:0051604	protein maturation	3	25	0.0463224989219289	Low SPEI / HTLP	GWC
PCO4/PCO5/RAV1	BP	GO:0070482	response to oxygen levels	3	25	0.0463224989219289	Low SPEI / HTLP	GWC
PILS5/EMS1/CR4	BP	GO:0090698	post-embryonic plant morphogenesis	3	25	0.0463224989219289	Low SPEI / HTLP	GWC
EMS1	BP	GO:0048656	anther wall tapetum formation	1	25	0.047791428625312	Low SPEI / HTLP	GWC
EMS1	BP	GO:0048657	anther wall tapetum cell differentiation	1	25	0.047791428625312	Low SPEI / HTLP	GWC
RBOHF	BP	GO:0002679	respiratory burst involved in defense response	1	25	0.0491387106791164	Low SPEI / HTLP	GWC
CR4	BP	GO:0032877	positive regulation of DNA endoreduplication	1	25	0.0491387106791164	Low SPEI / HTLP	GWC
CR4	BP	GO:2000105	positive regulation of DNA-templated DNA replication	1	25	0.0491387106791164	Low SPEI / HTLP	GWC
PCO4/PCO5	MF	GO:0016702	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	2	26	0.0341005157678652	Low SPEI / HTLP	GWC
PUB43/SBP1/ATL73/PUB23/UPL1/PRT6	MF	GO:0004842	ubiquitin-protein transferase activity	6	52	0.0254434633418956	High SPEI / LTHP	DWC
ROC1/ATL73/UPL2/BARD1/BRCA1/PRT6	MF	GO:0004842	ubiquitin-protein transferase activity	6	52	0.0091407743136777	Low SPEI / HTLP	DWC
ROC1/ATL73/BARD1/BRCA1	MF	GO:0004842	ubiquitin-protein transferase activity	4	26	0.0341005157678652	Low SPEI / HTLP	GWC
ROC2/ROC1	MF	GO:0016018	cyclosporin A binding	2	25	0.00939780570147978	High SPEI / LTHP	RWC

ROC2/ROC1	MF	GO:0016018	cyclosporin A binding	2	25	0.00939780570147978	High SPEI / LTHP	GWC
ROC1/ROC2	MF	GO:0016018	cyclosporin A binding	2	52	0.0091407743136777	Low SPEI / HTLP	DWC
ROC1/ROC2	MF	GO:0016018	cyclosporin A binding	2	13	0.00158444578096837	Low SPEI / HTLP	RWC
ROC1/ROC2	MF	GO:0016018	cyclosporin A binding	2	26	0.011758586133132	Low SPEI / HTLP	GWC
ROC2/ROC1	MF	GO:0003755	peptidyl-prolyl cis-trans isomerase activity	2	25	0.0476551726418193	High SPEI / LTHP	RWC
ROC2/ROC1	MF	GO:0003755	peptidyl-prolyl cis-trans isomerase activity	2	25	0.0476551726418193	High SPEI / LTHP	GWC
ROC1/ROC2	MF	GO:0003755	peptidyl-prolyl cis-trans isomerase activity	2	13	0.00819606676817156	Low SPEI / HTLP	RWC
ROC1/ROC2	MF	GO:0003755	peptidyl-prolyl cis-trans isomerase activity	2	26	0.0341005157678652	Low SPEI / HTLP	GWC
ROC2/ROC1	MF	GO:0016859	cis-trans isomerase activity	2	25	0.0476551726418193	High SPEI / LTHP	RWC
ROC2/ROC1	MF	GO:0016859	cis-trans isomerase activity	2	25	0.0476551726418193	High SPEI / LTHP	GWC
ROC1/ROC2	MF	GO:0016859	cis-trans isomerase activity	2	13	0.00819606676817156	Low SPEI / HTLP	RWC
CTF7/SRS1/MYB56/ATL72/ATL73/KAN2/SEU	BP	GO:0048438	floral whorl development	7	52	0.00885876135826332	Low SPEI / HTLP	DWC
ATL72/ATL73/EMS1/KAN2	BP	GO:0048438	floral whorl development	4	25	0.0133851206006658	Low SPEI / HTLP	GWC
ATL73/BARD1/BRCA1	BP	GO:0042304	regulation of fatty acid biosynthetic process	3	52	0.00885876135826332	Low SPEI / HTLP	DWC
ATL73/BARD1/BRCA1	BP	GO:0042304	regulation of fatty acid biosynthetic process	3	25	0.002210035276728	Low SPEI / HTLP	GWC
BARD1/BRCA1	BP	GO:1901874	negative regulation of post-translational protein modification	2	52	0.00885876135826332	Low SPEI / HTLP	DWC
BARD1/BRCA1	BP	GO:1901874	negative regulation of post-translational protein modification	2	25	0.002210035276728	Low SPEI / HTLP	GWC
BARD1/BRCA1	BP	GO:0035065	regulation of histone acetylation	2	52	0.00885876135826332	Low SPEI / HTLP	DWC
BARD1/BRCA1	BP	GO:0035065	regulation of histone acetylation	2	25	0.002210035276728	Low SPEI / HTLP	GWC
BARD1/BRCA1	BP	GO:0045922	negative regulation of fatty acid metabolic process	2	52	0.00885876135826332	Low SPEI / HTLP	DWC
BARD1/BRCA1	BP	GO:0045922	negative regulation of fatty acid metabolic process	2	25	0.002210035276728	Low SPEI / HTLP	GWC
BARD1/BRCA1	BP	GO:1901983	regulation of protein acetylation	2	52	0.00885876135826332	Low SPEI / HTLP	DWC
BARD1/BRCA1	BP	GO:1901983	regulation of protein acetylation	2	25	0.002210035276728	Low SPEI / HTLP	GWC
BARD1/BRCA1	BP	GO:2000756	regulation of peptidyl-lysine acetylation	2	52	0.00885876135826332	Low SPEI / HTLP	DWC
BARD1/BRCA1	BP	GO:2000756	regulation of peptidyl-lysine acetylation	2	25	0.002210035276728	Low SPEI / HTLP	GWC
ATL73/BARD1/BRCA1	BP	GO:0019217	regulation of fatty acid metabolic process	3	52	0.00885876135826332	Low SPEI / HTLP	DWC
ATL73/BARD1/BRCA1	BP	GO:0019217	regulation of fatty acid metabolic process	3	25	0.002210035276728	Low SPEI / HTLP	GWC
PCO4/PCO5/ATPK2/SEU/RAV1	BP	GO:0001666	response to hypoxia	5	52	0.0320433224645579	Low SPEI / HTLP	DWC
PCO4/PCO5/RAV1	BP	GO:0001666	response to hypoxia	3	25	0.0463224989219289	Low SPEI / HTLP	GWC

BARD1/BRCA1	BP	GO:0062014	negative regulation of small molecule metabolic process	2	52	0.0320811557366797	Low SPEI / HTLP	DWC
BARD1/BRCA1	BP	GO:0062014	negative regulation of small molecule metabolic process	2	25	0.00953400046325518	Low SPEI / HTLP	GWC
VPS2.2/VPS2.3	BP	GO:0032509	endosome transport via multivesicular body sorting pathway	2	52	0.0363703407570858	Low SPEI / HTLP	DWC
VPS2.2/VPS2.3	BP	GO:0032509	endosome transport via multivesicular body sorting pathway	2	25	0.0104211060491972	Low SPEI / HTLP	GWC
PCO4/PCO5/ROC1	BP	GO:0051606	detection of stimulus	3	52	0.0380395850491969	Low SPEI / HTLP	DWC
PCO4/PCO5/ROC1	BP	GO:0051606	detection of stimulus	3	25	0.00771466735267972	Low SPEI / HTLP	GWC
VPS2.2/VPS2.3	BP	GO:0071985	multivesicular body sorting pathway	2	52	0.0380395850491969	Low SPEI / HTLP	DWC
VPS2.2/VPS2.3	BP	GO:0071985	multivesicular body sorting pathway	2	25	0.0108635867688454	Low SPEI / HTLP	GWC
ATL73/BARD1/BRCA1/AAE13	BP	GO:0006633	fatty acid biosynthetic process	4	52	0.0399531726328381	Low SPEI / HTLP	DWC
ATL73/BARD1/BRCA1/AAE13	BP	GO:0006633	fatty acid biosynthetic process	4	25	0.00535012667331132	Low SPEI / HTLP	GWC
ATPK2/BARD1/BRCA1	BP	GO:0051247	positive regulation of protein metabolic process	3	52	0.0423876142630169	Low SPEI / HTLP	DWC
BARD1/BRCA1	BP	GO:0051247	positive regulation of protein metabolic process	2	25	0.0463224989219289	Low SPEI / HTLP	GWC
RING1A/BARD1/BRCA1	CC	GO:0000152	nuclear ubiquitin ligase complex	3	50	0.00786034645113005	Low SPEI / HTLP	DWC
TTM3/BARD1/BRCA1	CC	GO:0000152	nuclear ubiquitin ligase complex	3	25	0.00077226028620362 3	Low SPEI / HTLP	GWC
RING1A/ROC1/UPL2/BARD1/BRCA1/PRT6	CC	GO:0000151	ubiquitin ligase complex	6	50	0.00786034645113005	Low SPEI / HTLP	DWC
TTM3/ROC1/BARD1/BRCA1	CC	GO:0000151	ubiquitin ligase complex	4	25	0.00346626060289898	Low SPEI / HTLP	GWC
VPS2.2/VPS2.3	CC	GO:0000815	ESCRT III complex	2	50	0.00862574638997222	Low SPEI / HTLP	DWC
VPS2.2/VPS2.3	CC	GO:0000815	ESCRT III complex	2	25	0.0013334468254685	Low SPEI / HTLP	GWC

Table S8. Gene Ontology Analysis of genes corresponding to significant SNPs in association analysis of growth in term of SPEI value during the time-period June-August.

Genes	ONTOLOGY	GO ID	Gene Ontology term	Size overlap term / Count	Size overlap category	p.adjust	SPEI / Temperature and Precipitation	Association
ATG18C/ATG18D	CC	GO:0034045	phagophore assembly site membrane	2	48	0.027497619213474	High SPEI or LTHP	DWC
ATG18C/ATG18D/PCAP1	MF	GO:0080025	phosphatidylinositol-3,5-bisphosphate binding	3	49	0.00117701906024141	High SPEI or LTHP	DWC
INT1/INT2	MF	GO:0005365	myo-inositol transmembrane transporter activity	2	49	0.00936977293333363	High SPEI or LTHP	DWC
MAN1/MAN7	MF	GO:0016985	mannan endo-1,4-beta-mannosidase activity	2	49	0.00936977293333363	High SPEI or LTHP	DWC
MAN1/MAN7	MF	GO:0004567	beta-mannosidase activity	2	49	0.00936977293333363	High SPEI or LTHP	DWC
INT1/INT2	MF	GO:0015166	polyol transmembrane transporter activity	2	49	0.00936977293333363	High SPEI or LTHP	DWC
RPL5A/RPL5B	MF	GO:0008097	5S rRNA binding	2	49	0.0130574043802036	High SPEI or LTHP	DWC
ATG18C/ATG18D	MF	GO:0032266	phosphatidylinositol-3-phosphate binding	2	49	0.0147061487397375	High SPEI or LTHP	DWC
MIEL1/PUB62/ATL73/PRT6/BB	MF	GO:0004842	ubiquitin-protein transferase activity	5	49	0.0383056687317896	High SPEI or LTHP	DWC
AHL10/RAD51B	MF	GO:0000217	DNA secondary structure binding	2	49	0.0498965754024184	High SPEI or LTHP	DWC
GAD1/GAD3	BP	GO:0006538	glutamate catabolic process	2	20	0.00808194382547344	High SPEI or LTHP	RWC
GAD1/GAD3	BP	GO:0043649	dicarboxylic acid catabolic process	2	20	0.00808194382547344	High SPEI or LTHP	RWC
CSE/CINV1/MKK1	BP	GO:0042542	response to hydrogen peroxide	3	20	0.0127076717439428	High SPEI or LTHP	RWC
GAD1/GAD3	BP	GO:0006536	glutamate metabolic process	2	20	0.0197128877370343	High SPEI or LTHP	RWC
MKK1/MKK2	BP	GO:0000165	MAPK cascade	2	20	0.0197128877370343	High SPEI or LTHP	RWC
MKK1/MKK2	BP	GO:0009875	pollen-pistil interaction	2	20	0.0197128877370343	High SPEI or LTHP	RWC
MKK1/MKK2/MKK6	MF	GO:0004708	MAP kinase kinase activity	3	20	0.00015098548031664 7	High SPEI or LTHP	RWC
GAD1/GAD3/UXS4	MF	GO:0016831	carboxy-lyase activity	3	20	0.00112979242046737	High SPEI or LTHP	RWC
ACBP1/ACBP2	BP	GO:0010288	response to lead ion	2	28	0.0159866749816312	High SPEI or LTHP	GWC
RING1A/ACBP1/ZOP1/PRT6/GAUT11	BP	GO:0009845	seed germination	5	28	0.0159866749816312	High SPEI or LTHP	GWC
ACBP1/ACBP2	MF	GO:0000062	fatty-acyl-CoA binding	2	28	0.00208104538304689	High SPEI or LTHP	GWC

ACBP1/ACBP2	MF	GO:0120227	acyl-CoA binding	2	28	0.00208104538304689	High SPEI or LTHP	GWC
ACBP1/ACBP2	MF	GO:1901567	fatty acid derivative binding	2	28	0.00208104538304689	High SPEI or LTHP	GWC
JMJ30	BP	GO:0042060	wound healing	1	3	0.049506524947887	Low SPEI or HTLP	DWC
JMJ30	BP	GO:0009742	brassinosteroid mediated signaling pathway	1	3	0.049506524947887	Low SPEI or HTLP	DWC
JMJ30	BP	GO:0043401	steroid hormone mediated signaling pathway	1	3	0.049506524947887	Low SPEI or HTLP	DWC
JMJ30	BP	GO:0071383	cellular response to steroid hormone stimulus	1	3	0.049506524947887	Low SPEI or HTLP	DWC
JMJ30	BP	GO:0071367	cellular response to brassinosteroid stimulus	1	3	0.049506524947887	Low SPEI or HTLP	DWC
RPL34	CC	GO:0015934	large ribosomal subunit	1	3	0.0351246995084778	Low SPEI or HTLP	DWC
RPL34	MF	GO:0003735	structural constituent of ribosome	1	3	0.0354826869941103	Low SPEI or HTLP	DWC
PUB23	BP	GO:0002679	respiratory burst involved in defense response	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
DSP1	BP	GO:0034472	snRNA 3'-end processing	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
PUB23	BP	GO:0045730	respiratory burst	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
DSP1	BP	GO:0016180	snRNA processing	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
GH3.5	BP	GO:0051176	positive regulation of sulfur metabolic process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
GH3.5	BP	GO:0010120	camalexin biosynthetic process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
GH3.5	BP	GO:0052317	camalexin metabolic process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
DSP1	BP	GO:0016073	snRNA metabolic process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
KCR1	BP	GO:0042761	very long-chain fatty acid biosynthetic process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
PUB23	BP	GO:0002252	immune effector process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
GH3.5	BP	GO:0009403	toxin biosynthetic process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
GH3.5	BP	GO:0009700	indole phytoalexin biosynthetic process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
GH3.5	BP	GO:0046217	indole phytoalexin metabolic process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
GH3.5	BP	GO:0052314	phytoalexin metabolic process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
GH3.5	BP	GO:0052315	phytoalexin biosynthetic process	1	7	0.0296830827548597	Low SPEI or HTLP	RWC
MYB26	BP	GO:0009901	anther dehiscence	1	7	0.0297141449403069	Low SPEI or HTLP	RWC
MYB26	BP	GO:0009900	dehiscence	1	7	0.0387390505491611	Low SPEI or HTLP	RWC
GH3.5	BP	GO:1900376	regulation of secondary metabolite biosynthetic process	1	7	0.0387390505491611	Low SPEI or HTLP	RWC

KCR1	BP	GO:0000038	very long-chain fatty acid metabolic process	1	7	0.0417914377961174	Low SPEI or HTLP	RWC
DSP1	BP	GO:0031123	RNA 3'-end processing	1	7	0.0492692988126216	Low SPEI or HTLP	RWC
GH3.5	MF	GO:0010279	indole-3-acetic acid amido synthetase activity	1	6	0.0274398990190381	Low SPEI or HTLP	RWC
RBP47C'	MF	GO:0008143	poly(A) binding	1	6	0.0274398990190381	Low SPEI or HTLP	RWC
RBP47C'	MF	GO:0070717	poly-purine tract binding	1	6	0.0274398990190381	Low SPEI or HTLP	RWC
TIR	MF	GO:0050135	NAD(P)+ nucleosidase activity	1	6	0.0274398990190381	Low SPEI or HTLP	RWC
TIR	MF	GO:0003953	NAD+ nucleosidase activity	1	6	0.0274398990190381	Low SPEI or HTLP	RWC
TIR	MF	GO:0061809	NAD+ nucleotidase, cyclic ADP-ribose generating	1	6	0.0274398990190381	Low SPEI or HTLP	RWC
SYP41/SYP43	BP	GO:0043001	Golgi to plasma membrane protein transport	2	4	6.72834354172023e-05	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0061951	establishment of protein localization to plasma membrane	2	4	6.72834354172023e-05	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0072659	protein localization to plasma membrane	2	4	6.72834354172023e-05	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:1990778	protein localization to cell periphery	2	4	6.72834354172023e-05	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0006893	Golgi to plasma membrane transport	2	4	9.68505796772033e-05	Low SPEI or HTLP	GWC
SYP41/SYP43/JMJ30	BP	GO:0071407	cellular response to organic cyclic compound	3	4	0.00021497317085675 8	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0007030	Golgi organization	2	4	0.00021497317085675 8	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0006896	Golgi to vacuole transport	2	4	0.00024313176375718 5	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0048278	vesicle docking	2	4	0.00055140376744399 7	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0022406	membrane docking	2	4	0.00060669599674878 4	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0006906	vesicle fusion	2	4	0.00062711912617501 6	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0051668	localization within membrane	2	4	0.00079907686974608 5	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0061025	membrane fusion	2	4	0.00079907686974608 5	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0048284	organelle fusion	2	4	0.0010801505903157	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0007034	vacuolar transport	2	4	0.00131143529127551	Low SPEI or HTLP	GWC
SYP41/SYP43	BP	GO:0016050	vesicle organization	2	4	0.00131143529127551	Low SPEI or HTLP	GWC

SYP41/SYP43	BP	GO:0009658	chloroplast organization	2	4	0.00449504799701774	Low SPEI or HTLP	GWC
SYP 43	BP	GO:0009306	protein secretion	1	4	0.0187077401848535	Low SPEI or HTLP	GWC
SYP 43	BP	GO:0035592	establishment of protein localization to extracellular region	1	4	0.0187077401848535	Low SPEI or HTLP	GWC
SYP 43	BP	GO:0071692	protein localization to extracellular region	1	4	0.0187077401848535	Low SPEI or HTLP	GWC
SYP 43	BP	GO:1900150	regulation of defense response to fungus	1	4	0.0210723510219699	Low SPEI or HTLP	GWC
SYP 43	BP	GO:0032940	secretion by cell	1	4	0.0405117770854127	Low SPEI or HTLP	GWC
SYP 43	BP	GO:0046903	secretion	1	4	0.0435673954306648	Low SPEI or HTLP	GWC
JMJ30	BP	GO:2000028	regulation of photoperiodism, flowering	1	4	0.0447769791108422	Low SPEI or HTLP	GWC
JMJ30	BP	GO:1900140	regulation of seedling development	1	4	0.0488778970870973	Low SPEI or HTLP	GWC
SYP41/SYP43	CC	GO:0031201	SNARE complex	2	4	0.00167112237131154	Low SPEI or HTLP	GWC
SYP41/SYP43	CC	GO:0005802	trans-Golgi network	2	4	0.00973775300169354	Low SPEI or HTLP	GWC
SYP 41	CC	GO:0030140	trans-Golgi network transport vesicle	1	4	0.0197249854039083	Low SPEI or HTLP	GWC
SYP41/SYP43	MF	GO:0005484	SNAP receptor activity	2	4	0.00332763196755325	Low SPEI or HTLP	GWC
SYP41/SYP43	MF	GO:0000149	SNARE binding	2	4	0.00332763196755325	Low SPEI or HTLP	GWC
MBF1A	MF	GO:0003713	transcription coactivator activity	1	4	0.0222683326175012	Low SPEI or HTLP	GWC
RPL5/RPL5A/RPL 5B	CC	GO:0022625	cytosolic large ribosomal subunit	3	48	0.027497619213474	High SPEI or LTHP	DWC
RPL34	CC	GO:0022625	cytosolic large ribosomal subunit	1	3	0.0351246995084778	Low SPEI or HTLP	DWC
JMJ30	BP	GO:1900457	regulation of brassinosteroid mediated signaling pathway	1	3	0.049506524947887	Low SPEI or HTLP	DWC
JMJ30	BP	GO:1900457	regulation of brassinosteroid mediated signaling pathway	1	4	0.0130142872181786	Low SPEI or HTLP	GWC
JMJ30	BP	GO:1990110	callus formation	1	3	0.049506524947887	Low SPEI or HTLP	DWC
JMJ30	BP	GO:1990110	callus formation	1	4	0.0187077401848535	Low SPEI or HTLP	GWC
JMJ30	BP	GO:0080022	primary root development	1	3	0.049506524947887	Low SPEI or HTLP	DWC
JMJ30	BP	GO:0080022	primary root development	1	4	0.0227080250107758	Low SPEI or HTLP	GWC
JMJ30	BP	GO:0042752	regulation of circadian rhythm	1	3	0.049506524947887	Low SPEI or HTLP	DWC
JMJ30	BP	GO:0042752	regulation of circadian rhythm	1	4	0.0354816609793298	Low SPEI or HTLP	GWC
JMJ30	BP	GO:0048545	response to steroid hormone	1	3	0.049506524947887	Low SPEI or HTLP	DWC
JMJ30	BP	GO:0048545	response to steroid hormone	1	4	0.037236411179565	Low SPEI or HTLP	GWC

JMJ30	BP	GO:0045814	negative regulation of gene expression, epigenetic	1	3	0.049506524947887	Low SPEI or HTLP	DWC
JMJ30	BP	GO:0045814	negative regulation of gene expression, epigenetic	1	4	0.0377667013919315	Low SPEI or HTLP	GWC
JMJ30	CC	GO:0000791	euchromatin	1	3	0.0256906222268034	Low SPEI or HTLP	DWC
JMJ30	CC	GO:0000791	euchromatin	1	4	0.0142652169748869	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0046975	histone H3K36 methyltransferase activity	1	3	0.0173043590524825	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0046975	histone H3K36 methyltransferase activity	1	4	0.016239160088453	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0141052	histone H3 demethylase activity	1	3	0.0173043590524825	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0141052	histone H3 demethylase activity	1	4	0.016239160088453	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0032452	histone demethylase activity	1	3	0.0173043590524825	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0032452	histone demethylase activity	1	4	0.016239160088453	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0140457	protein demethylase activity	1	3	0.0173043590524825	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0140457	protein demethylase activity	1	4	0.016239160088453	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0035064	methylated histone binding	1	3	0.0173043590524825	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0035064	methylated histone binding	1	4	0.016239160088453	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0140034	methylation-dependent protein binding	1	3	0.0173043590524825	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0140034	methylation-dependent protein binding	1	4	0.016239160088453	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0032451	demethylase activity	1	3	0.0183120689022238	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0032451	demethylase activity	1	4	0.0202292283048687	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0140938	histone H3 methyltransferase activity	1	3	0.0183120689022238	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0140938	histone H3 methyltransferase activity	1	4	0.0209887797096461	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0016706	2-oxoglutarate-dependent dioxygenase activity	1	3	0.0269748063060552	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0016706	2-oxoglutarate-dependent dioxygenase activity	1	4	0.0307622934869159	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0042393	histone binding	1	3	0.0349955686191481	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0042393	histone binding	1	4	0.0403116703690184	Low SPEI or HTLP	GWC
JMJ30	MF	GO:0140993	histone modifying activity	1	3	0.039283951789194	Low SPEI or HTLP	DWC
JMJ30	MF	GO:0140993	histone modifying activity	1	4	0.0484137582303306	Low SPEI or HTLP	GWC

Table S9. Sequencing information for the 29 RRBS samples.

Sample	Reads	Bases	Q20	Q20(%)	Q30	Q30(%)	GC (%)	% coverage	BS- conversion
FURUBM104	219778377	6,59E+10	6,39E+10	96,89	6,08E+10	92,18	30	3	99,75
FURUBM108	209150582	6,27E+10	6,03E+10	96,07	5,70E+10	90,9	32	2,85	99,32
FURUBM35	205754747	6,17E+10	5,98E+10	96,81	5,67E+10	91,82	28	2,81	99,83
FURUBM43	278885563	8,37E+10	8,05E+10	96,26	7,62E+10	91,04	30	3,8	99,83
FURUBM56	156157900	4,68E+10	4,54E+10	96,85	4,32E+10	92,23	37	2,13	99,03
FURUBM66	211611899	6,35E+10	6,14E+10	96,65	5,81E+10	91,58	29	2,89	99,35
FURUBM77	268999490	8,07E+10	7,72E+10	95,61	7,25E+10	89,8	33	3,67	99,77
FURUBM82	304770247	4,55E+10	4,38E+10	96,45	4,15E+10	91,36	26	2,07	99,89
FURUBM83	220210327	6,61E+10	6,38E+10	96,6	6,05E+10	91,62	29	3	99,44
FURUBS22	247415516	7,42E+10	7,22E+10	97,27	6,90E+10	92,92	28	3,37	99,75
FURUBS45	201133026	6,03E+10	5,86E+10	97,18	5,58E+10	92,53	29	2,74	99,31
FURUBS87	228603400	6,86E+10	6,69E+10	97,49	6,40E+10	93,31	27	3,12	99,85
FURUG108	267747275	8,03E+10	7,76E+10	96,67	7,36E+10	91,61	28	3,65	99,88
FURUG22	282789645	8,48E+10	8,11E+10	95,57	7,60E+10	89,64	31	3,86	99,88
FURUG24	230188025	6,91E+10	6,68E+10	96,76	6,36E+10	92,05	28	3,14	99,86
FURUG25	237499478	7,12E+10	6,91E+10	96,94	6,56E+10	92,08	28	3,24	99,38
FURUG42	257213561	7,72E+10	7,41E+10	96,07	7,01E+10	90,9	31	3,51	99,81
FURUG43	272274729	8,17E+10	7,80E+10	95,53	7,31E+10	89,48	32	3,71	99,78
FURUG73	234612648	7,04E+10	6,81E+10	96,7	6,45E+10	91,57	30	3,2	99,91
FURUG78	226306247	6,79E+10	6,57E+10	96,83	6,23E+10	91,76	27	3,09	99,49
FURUG86	279432855	8,38E+10	8,08E+10	96,39	7,64E+10	91,09	29	3,81	99,59
FURUG87	216773103	6,50E+10	6,30E+10	96,86	5,98E+10	91,96	29	2,96	99,87
FURUG92	306136510	4,58E+10	4,41E+10	96,46	4,17E+10	91,17	24	2,08	99,1
FURUS23	206946745	6,21E+10	6,01E+10	96,81	5,70E+10	91,78	28	2,82	99,87
FURUS34	328683713	9,86E+10	9,43E+10	95,58	8,83E+10	89,55	31	4,48	99,82
FURUS55	323593471	4,84E+10	4,66E+10	96,42	4,41E+10	91,11	25	2,2	99,89
FURUS75	222081252	6,66E+10	6,42E+10	96,3	6,06E+10	90,92	27	3,03	99,87
FURUS82	316780240	9,50E+10	9,07E+10	95,39	8,49E+10	89,34	32	4,32	99,83
FURUS95	292089982	8,76E+10	8,46E+10	96,51	8,03E+10	91,66	30	3.98	99,89

Table S10. BLASTn of NvS showing top 10 hits of differentially methylated fragments for read 1 and 2 together.

Query	meth.sites	meth.meth_hi	Number of HSPs	Lowest E-value	Accession
CpG					
dedRef_2315294-2315295_8	1	S	27232	2,01E-58	ChrUn38
dedRef_755097-755098_9	1	S	64	9,77E-50	ChrUn62
dedRef_1054162-1054163_53	3	N	2	5,06E-47	scaffold_698
dedRef_2173022-2173022_10	3	N	6251	5,11E-47	ChrUn01
dedRef_2907510-2907511_16-2	5	S	1	7,38E-45	scaffold_5673
dedRef_4857208-4857210_5	2	S	1	3,93E-42	scaffold_27319
dedRef_253744-253745_43	2	N	3	1,29E-41	ChrUn45
dedRef_664348-664349_46	1	S	1	4,74E-41	ChrUn70
dedRef_1142677-1142678_31	3	S	1	5,78E-40	scaffold_63823
dedRef_1005891-1005892_55	2	N	1	5,78E-40	scaffold_2506
CHG					
dedRef_90022-90023_4	2	N	3	6,98E-39	scaffold_12304
dedRef_255777-255778_43	3	N	1	1,48E-34	scaffold_41156
dedRef_1049112-1049113_23	2	S	23	6,59E-33	ChrUn47
dedRef_973251-973252_36	3	S	48	2,8E-31	ChrUn23
dedRef_676286-676287_33	3	S	90	9,78E-31	ChrUn75
dedRef_614715-614716_35	2	S	1	4,01E-29	scaffold_54049
dedRef_772016-772016_23	3	N	1	4,01E-29	scaffold_86256
dedRef_526259-526259_27	2	S	1	4,16E-29	scaffold_49835
dedRef_1212718-1212719_31	3	N	1	1,41E-28	scaffold_75899
dedRef_236188-236189_34	3	N	10	1,45E-28	ChrUn80
CHH					
dedRef_1632067-1632068_34	30	N	1	7,51E-45	scaffold_83203
dedRef_4715855-4715858_13-62	1	S	2	1,37E-41	scaffold_1498
dedRef_4581053-4581054_11	1	N	6	6,98E-39	scaffold_12778
dedRef_647431-647432_17	15	S	5	8,5E-38	scaffold_3219
dedRef_545611-545612_7	13	S	2	1,26E-35	scaffold_21629
dedRef_3889593-3889594_10	1	N	1	1,89E-33	scaffold_27684
dedRef_639626-639627_22	16	N	7	6,59E-33	scaffold_126
dedRef_22713-22714_4	1	S	15	2,46E-31	ChrUn18
dedRef_1923928-1923929_6	1	S	5	1,26E-28	scaffold_41490
dedRef_21888-21889_27	14	N	5	1,75E-27	ChrUn169

Table S11. BLASTn of NvS showing top 10 hits of differentially methylated motifs for read 1 & read 2 together.

meth.ID	Query	Number of HSPs	Lowest E-value	Accession
		CpG		
536953	dedRef_948667-948668_37-1	1	3,1945E-43	scaffold_13857
360543	dedRef_1005891-1005892_55	1	5,7757E-40	scaffold_2506
472183	dedRef_351881-351882_5-1	6	1,609E-33	ChrUn68
395681	dedRef_1081898-1081899_35	31	1,8881E-33	ChrUn30
867534	dedRef_2245176-2245177_4	2	1,9045E-33	scaffold_11619
78352	dedRef_938801-938802_41-1	3	6,5901E-33	scaffold_2915
84207	dedRef_824291-824292_30-1	35	7,8887E-32	ChrUn109
76196	dedRef_969876-969877_25-1	1	8,0283E-32	scaffold_41037
144456	dedRef_965818-965819_46-2	1	9,7805E-31	scaffold_69642
194514	dedRef_1378288-1378289_35-1	1	3,4137E-30	scaffold_60464
		CHG		
365974	dedRef_2226864-2226867_8-1	34	8,5719E-38	ChrUn66
1130248	dedRef_2068626-2068631_11	1	8,2634E-37	scaffold_4355
271908	dedRef_255777-255778_43	1	1,4825E-34	scaffold_41156
313804	dedRef_1209028-1209029_30	1	3,2653E-30	scaffold_45259
33714	dedRef_907955-907956_37-4	171	1,1915E-29	ChrUn37
330141	dedRef_1557051-1557052_36	3	1,1915E-29	scaffold_34497
313951	dedRef_1212718-1212719_31	1	1,4137E-28	scaffold_75899
169318	dedRef_292153-292154_38	4	1,4516E-28	ChrUn61
1100894	dedRef_343903-343904_29	1	1,4516E-28	scaffold_4294
1492355	dedRef_697818-697819_39	9	1,4516E-28	ChrUn32
		CHH		
1468074/1311832	dedRef_899225-899226_33	2	3,61E-55	ChrUn103
2845051	dedRef_976008-976009_37	1	1,6548E-40	scaffold_12988
2891618	dedRef_756654-756655_36	3	5,7757E-40	scaffold_3504
11773566	dedRef_141417-141418_37	1	5,4095E-34	scaffold_75233
599291	dedRef_790086-790087_32-2	1	6,5328E-33	scaffold_29057
389986	dedRef_925549-925550_31	1	2,3002E-32	ChrUn222
586701	dedRef_676254-676256_35-14	1	3,4137E-30	scaffold_60937
1007820	dedRef_1217920-1217921_36-8	33	3,4137E-30	ChrUn108
1134599	dedRef_1349776-1349777_34	5	3,4434E-30	ChrUn156
1134599	dedRef_1109816-1109817_31	2	1,1708E-29	ChrUn40

Table S12. BLASTn of TvS showing top 10 hits of differentially methylated fragments for read 1 and read 2 together.

Query	meth.sites	meth.meth_hi	Number of HSPs	Lowest E-value	Accession
Read 1 - CpG					
dedRef_1529438-1529439_10	4	Tall	2	2,1356E-26	scaffold_12478
dedRef_319768-319769_37	4	Tall	2	5,4142E-15	scaffold_2422
dedRef_2018734-2018736_2	1	Short	1	0,00039324	scaffold_62495
Read 2 - CpG					
dedRef_2173022-2173022_10	3	Tall	6251	5,106E-47	ChrUn01
dedRef_4857208-4857210_5	2	Short	1	3,9255E-42	scaffold_27319
dedRef_2869945-2869946_7-1	3	Tall	1	2,0159E-39	scaffold_1769
dedRef_4823635-4823636_6	3	Tall	26	2,0335E-39	ChrUn186
dedRef_623102-623103_46	5	Tall	14	2,4345E-38	ChrUn172
dedRef_823610-823611_38	4	Short	1	2,9919E-37	scaffold_24384
dedRef_208019-208020_47	4	Tall	11	1,2169E-35	ChrUn106
dedRef_255757-255758_27	3	Tall	26	1,2722E-35	ChrUn13
dedRef_282377-282379_28	1	Short	1	1,2722E-35	scaffold_37791
dedRef_1556511-1556515_33	3	Short	2	1,5229E-34	scaffold_37859
Read 1 - CHG					
dedRef_334714-334715_38	2	Short	1	5,0179E-47	scaffold_58740
dedRef_1165881-1165882_42	2	Tall	7	7,5126E-45	ChrUn60
dedRef_1294492-1294493_3-1	1	Short	11	2,6222E-44	ChrUn32
dedRef_1107886-1107887_48	2	Tall	2	9,8071E-37	scaffold_34174
dedRef_1379316-1379317_1	1	Tall	2	3,6449E-36	scaffold_15634
dedRef_268993-268994_31	3	Short	121	1,8717E-33	ChrUn54
dedRef_3645364-3645365_1	1	Tall	1	5,0428E-33	scaffold_46177
dedRef_250037-250038_30	2	Short	1	6,5328E-33	scaffold_81503
dedRef_3109089-3109090_7	3	Short	684	2,2002E-32	ChrUn18
dedRef_837175-837176_27	1	Short	1	2,3002E-32	scaffold_3927
Read 2 - CHG					
dedRef_3069618-3069619_3	2	Tall	97	1,9028E-52	ChrUn31
dedRef_1577893-1577894_33-3	1	Short	1	4,1551E-48	scaffold_54523
dedRef_1541112-1541113_36	3	Short	1	5,4095E-34	scaffold_65379
dedRef_1049112-1049113_23	2	Tall	23	6,5901E-33	ChrUn47
dedRef_1010182-1010183_35	2	Short	1	2,2202E-32	scaffold_96683
dedRef_1056635-1056636_31	3	Tall	1	7,9585E-32	scaffold_45023

dedRef_973251-973252_36	3	Short	48	2,8022E-31	ChrUn23
dedRef_536389-536389_39	2	Tall	1	3,4137E-30	scaffold_36806
dedRef_614715-614716_35	3	Short	1	4,0141E-29	scaffold_54049
dedRef_1212718-1212719_31	3	Short	1	1,4137E-28	scaffold_75899
Read 1 - CHH					
dedRef_105983-105984_3	1	Tall	30	1,7668E-46	ChrUn04
dedRef_1506865-1506869_10	2	Tall	12	9,1522E-44	ChrUn16
dedRef_1334186-1334187_32	1	Tall	1	1,6548E-40	scaffold_1054
dedRef_560523-560524_35	1	Short	11	1,0352E-36	ChrUn28
dedRef_4150913-4150914_14	1	Short	1	1,5498E-34	scaffold_9690
dedRef_1452098-1452099_34	1	Tall	4	7,9585E-32	scaffold_515
dedRef_22713-22714_4	1	Short	15	2,461E-31	ChrUn18
dedRef_4979461-4979462_11	1	Short	10	9,8656E-31	ChrUn156
dedRef_875349-875350_10	2	Short	2	2,1356E-26	scaffold_83111
dedRef_3576178-3576179_2	3	Tall	2	2,1452E-25	scaffold_43898
Read 2 - CHH					
dedRef_1436209-1436210_36	20	Tall	306	1,0443E-36	ChrUn42
dedRef_802343-802344_32	27	Tall	3	2,8022E-31	scaffold_4186
dedRef_41277-41280_29-4	13	Tall	2	1,4263E-28	scaffold_17687
dedRef_334942-334943_37	29	Tall	5	7,3885E-26	scaffold_10240
dedRef_1428971-1428972_20	10	Tall	3	7,5193E-26	scaffold_32343
dedRef_1455166-1455167_23	11	Tall	3	7,5193E-26	scaffold_16569
dedRef_1247741-1247742_13-2	13	Short	1085	2,6245E-25	ChrUn107
dedRef_335913-335914_23	17	Tall	1	3,1139E-24	scaffold_34480
dedRef_1550897-1550898_28	15	Tall	1	1,116E-23	scaffold_25985
dedRef_762892-762893_25	24	Tall	2	1,3595E-22	scaffold_95739

Table S13. BLASTn of TvS showing top 10 hits of differentially methylated motifs for read 1 and read 2 together.

meth.ID	Query	Number of HSPs	Lowest E-value	Accession
CpG				
267161	dedRef_1855544-1855544_0	2	2,1524E-45	scaffold_66244
272558	dedRef_293145-293146_44-4	80	7,51262E-45	ChrUn143
2500703	dedRef_2904326-2904327_9	1	8,91349E-44	scaffold_69079
536953	dedRef_948667-948668_37-1	1	3,19445E-43	scaffold_13857
1114383	dedRef_3436919-3436920_18	7	3,89163E-42	ChrUn03

1005069	dedRef_293665-293666_27	9	6,30202E-39	ChrUn168
1499785	dedRef_68373-68374_46	21	1,04427E-36	ChrUn19
1236594	dedRef_987810-987811_35	1	3,61317E-36	scaffold_30690
1480823	dedRef_612144-612145_43	3	4,32451E-35	ChrUn166
1472188	dedRef_58487-58488_46	11	1,54983E-34	ChrUn75
		CHG		
402829	dedRef_1167033-1167034_4	9	8,56432E-57	ChrUn89
2075369	dedRef_741894-741894_6	4	5,40468E-53	ChrUn166
832050	dedRef_130428-130429_42	177	9,68691E-50	ChrUn27
338557	dedRef_1852002-1852003_3-1	2	1,16976E-48	scaffold_25904
129721	dedRef_1577893-1577894_33-3	1	4,15511E-48	scaffold_54523
2133704	dedRef_820087-820088_37	6	1,29894E-47	ChrUn16
1903999	dedRef_504740-504741_7	13	1,33678E-47	ChrUn89
26293	dedRef_3153216-3153217_11-1	19	4,70983E-47	ChrUn63
876352	dedRef_3323316-3323317_8-3	1	6,16673E-46	scaffold_76113
87998	dedRef_1402718-1402719_38	2	6,16673E-46	scaffold_13619
		CHH		
839294	dedRef_899225-899226_33	2	3,60998E-55	ChrUn103
842329	dedRef_960473-960474_67-2	54	6,11311E-46	scaffold_78080
3603314	dedRef_4150913-4150914_14	1	1,54983E-34	scaffold_9690
506987	dedRef_141417-141418_37	1	5,40945E-34	scaffold_75233
3935573	dedRef_988170-988171_31	3	6,59006E-33	ChrUn121
724445	dedRef_1349776-1349777_34	5	3,44342E-30	ChrUn156
353336	dedRef_1109816-1109817_31	2	1,17079E-29	ChrUn40
814034	dedRef_1485964-1485965_44-2	2	4,12262E-29	scaffold_42180
6811852	dedRef_271941-271942_40	18	4,15878E-29	ChrUn81
1329428	dedRef_1211719-1211720_22-2	219	4,93427E-28	ChrUn49

Table S14. BLASTx of NvS showing top 10 hits of differentially methylated fragments for read 1 and 2 together.

Query	meth.sites	meth.meth_hi	Number of HSPs	Lowest E-value	Accession
		CpG			
dedRef_426340-426343_31-1	1	N	5	3,8645E-26	WP_206682097
dedRef_25086-25087_42	4	N	5	5,4959E-25	XP_036010831
dedRef_817851-817852_40	1	N	5	1,0549E-23	WP_302758310
dedRef_172015-172016_39-3	3	N	5	2,026E-23	WP_214284958

dedRef_1649087-1649089_45	16	N	5	1,0283E-22	NP_775626
dedRef_1112190-1112190_24-2	1	S	5	1,051E-22	WP_334412441
dedRef_1329529-1329530_47	8	S	5	6,0515E-22	WP_029550487
dedRef_412516-412517_25	2	S	5	1,3905E-21	WP_228029457
dedRef_901472-901473_28	1	S	5	2,0371E-21	WP_227649065
dedRef_1504774-1504775_29	2	N	5	3,0851E-21	WP_217806349
CHG					
dedRef_179953-179954_18	1	S	10	2,7001E-20	WP_206338328
dedRef_18060-18060_20	4	S	5	4,5839E-20	YP_003541050
dedRef_1003498-1003499_34	14	S	6	9,6109E-17	WP_303273946
dedRef_796422-796422_44	4	S	19	2,4789E-13	WP_271019098
dedRef_973251-973252_36	3	S	5	9,7943E-10	XP_020262272
dedRef_772423-772424_30	2	N	5	1,3734E-08	YP_009388296
dedRef_1501133-1501133_29	3	N	5	2,3297E-08	YP_009388272
dedRef_582116-582117_24	3	N	5	1,5257E-07	WP_214280456
dedRef_526259-526259_27	2	S	5	2,001E-07	YP_003934484
dedRef_1603247-1603249_31	2	N	5	2,593E-07	WP_111034173
CHH					
dedRef_883412-883413_50	41	S	5	1,1002E-27	WP_192529997
dedRef_132239-132239_45-3	51	S	6	2,8399E-16	XP_054110662
dedRef_347674-347675_31	1	S	5	4,1815E-12	XP_048322314
dedRef_860255-860255_34	52	N	5	7,4317E-12	WP_236882086
dedRef_1371339-1371340_31	16	N	5	4,0462E-07	WP_254062628
dedRef_1362256-1362257_31	13	S	5	1,1016E-06	WP_214280456
dedRef_639723-639724_26	14	N	5	1,1664E-06	WP_294211963
dedRef_785307-785308_25	14	N	5	1,6454E-06	WP_214280456
dedRef_1591560-1591561_26	14	N	5	2,3447E-06	WP_214280456
dedRef_723056-723057_22	11	N	5	2,9407E-06	WP_214280456

Table S15. BLASTx of NvS showing differentially methylated motifs for read 1 and read 2 together.

meth.ID	Query	Number of HSPs	Lowest E-value	Accession
Read 1 - CpG				
665556	dedRef_1473565-1473566_32	5	1,4065E-26	WP_305238619
1109014	dedRef_3416496-3416497_47	5	5,0067E-24	WP_106376840
714546/714550/714543/	dedRef_1649087-1649089_45	5	1,0283E-22	NP_775626

714535/714537/714538/				
714539/714540/714541/				
714542				
286912	dedRef_848521-848522_36	5	3,5988E-19	XP_058285325
530127	dedRef_776352-776353_60	5	3,9406E-14	XP_047302963
294448	dedRef_1816149-1816150_28-1	5	2,1076E-13	XP_054532921
330795	dedRef_1430779-1430780_18	35	6,4791E-12	WP_275218940
2581194	dedRef_3531305-3531306_47	5	1,8944E-09	XP_040845861
533885	dedRef_855587-855589_33	17	7,6873E-09	WP_208325343
358557	dedRef_3376029-3376030_57	3	3,0199E-08	XP_047302963
Read 2 - CpG				
249577	dedRef_4994647-4994647_37	5	2,1182E-26	WP_143768019
104725	dedRef_426340-426343_31-1	5	3,8645E-26	WP_206682097
763502	dedRef_25086-25087_42	5	5,4959E-25	XP_036010831
23764	dedRef_805625-805625_33-4	5	9,5514E-25	XP_054309898
294640	dedRef_2014210-2014210_16	5	2,5718E-21	WP_205923115
18063	dedRef_3003772-3003773_3-7	6	2,9027E-20	XP_063450681
352260	dedRef_865278-865278_51	11	1,0951E-19	WP_316486000
27464	dedRef_2608414-2608415_37-1	6	2,3777E-19	XP_063468535
165902	dedRef_459298-459299_32-1	5	2,8144E-17	WP_218012903
76184	dedRef_961322-961323_46-5	5	5,2744E-17	XP_055219682
Read 1 - CHG				
480487/480473/480475/				
480477	dedRef_1020880-1020883_24	10	1,0852E-25	WP_335896943
1462295	dedRef_3416496-3416497_47	5	5,0067E-24	WP_106376840
382486	dedRef_411196-411197_26	5	4,6113E-23	WP_198670135
1017209	dedRef_1700526-1700528_33	5	2,172E-13	XP_040584748
646382	dedRef_796422-796422_44	19	2,4789E-13	WP_271019098
232310	dedRef_175309-175310_40	5	3,1347E-13	XP_017744634
382985	dedRef_416648-416648_44	6	5,4318E-13	XP_011514698
365974	dedRef_2226864-2226867_8-1	1	2,34528	WP_149265971
Read 2 - CHG				
147397	dedRef_3336423-3336424_31-2	5	8,5012E-28	WP_220085038
191761	dedRef_720424-720425_20-3	5	8,4949E-24	WP_223009790
215993	dedRef_887106-887106_34-1	5	6,7395E-14	WP_236882086
169318	dedRef_292153-292154_38	5	3,6479E-10	XP_005822623

214910	dedRef_772423-772424_30	5	1,3734E-08	YP_009388296
70398	dedRef_906030-906031_23-1	5	1,3693E-07	WP_214280456
124226	dedRef_1517149-1517152_32	5	1,5931E-07	WP_214280456
332465	dedRef_1603247-1603249_31	5	2,593E-07	WP_111034173
203385	dedRef_241954-241955_34-1	5	3,6736E-07	WP_267994702
340976	dedRef_197472-197473_30	5	4,3418E-07	WP_254602374
Read 1 - CHH				
568802	dedRef_382319-382320_27-1	5	3,2166E-27	WP_157976345
497409	dedRef_733537-733537_34	5	2,7336E-26	WP_142319120
1027015	dedRef_1020880-1020883_24	10	1,0852E-25	WP_335896943
680955	dedRef_2514724-2514724_35-1	5	2,674E-24	WP_249080693
3163960/3163961/3163959/				
3163962	dedRef_3416496-3416497_47	5	5,0067E-24	WP_106376840
970524	dedRef_4817575-4817576_28	5	6,7637E-24	WP_258304901
839254/839294	dedRef_899225-899226_33	5	7,0827E-24	XP_047953659
1090459	dedRef_1562593-1562593_33-2	5	8,5418E-24	WP_220152734
623359/623335/623336/				
623354/623333	dedRef_346773-346774_36	5	1,0978E-23	WP_323563410
1232605/1232625	dedRef_36081-36081_23-2	5	4,5829E-23	WP_146032260
Read 2 - CHH				
1142826	dedRef_1523346-1523347_36-1	5	7,1117E-27	WP_283748784
1090994	dedRef_879036-879037_37	5	2,4715E-26	WP_330949972
1457815	dedRef_66709-66712_34	5	8,1703E-25	WP_257205104
1026519	dedRef_2107264-2107265_17-3	5	3,22E-22	WP_141021492
9877898	dedRef_3993608-3993608_27	5	5,1135E-22	WP_301161387
1079144	dedRef_693208-693209_37-1	5	7,3829E-22	WP_002874006
1085615	dedRef_801010-801011_53	78	5,7736E-21	WP_305212016
91276	dedRef_4944954-4944955_13-3	58	1,1487E-18	WP_255054528
10679562	dedRef_4864839-4864839_13	5	7,0298E-18	WP_208399076
2654711	dedRef_474407-474409_30	6	9,699E-18	WP_206311507

Table S16. BLASTx of TvS showing top 10 hits of differentially methylated fragments for read 1 and read 2 together.

Query	meth. sites	meth. meth_ hi	Number of HSPs	Lowest E-value	Accession	Description (E-value)
			Read 1 - CpG			
dedRef_1649087-1649089_45	16	Tall	5	1,02057E-22	NP_775626	pseudouridylate synthase RPUSD2 [Mus musculus]
dedRef_1648079-1648080_21	9	Short	40	2,50208E-19	XP_047298760	extensin-like [Homo sapiens] >ref|XP_047298761.1| extensin-like [Homo sapiens] >ref|XP_047298762.1| extensin-like [Homo sapiens]
dedRef_800731-800731_52	4	Tall	59	6,62922E-16	WP_259600334	hypothetical protein, partial [Burkholderia thailandensis]
			Read 2 - CpG			
dedRef_1545141-1545141_29	2	Short	5	8,11936E-28	WP_207567937	endonuclease/exonuclease/phosphatase family protein, partial [Cronobacter sakazakii]
dedRef_592472-592473_24-2	1	Tall	5	3,54782E-24	WP_323567329	hypothetical protein, partial [Pseudomonas sp. CCI2.4]
dedRef_1475486-1475486_32	4	Short	5	6,17721E-24	WP_164646752	hypothetical protein, partial [Pseudomonas viridiflava]
dedRef_411196-411197_26	2	Tall	5	4,57678E-23	WP_198670135	reverse transcriptase domain-containing protein, partial [Salmonella enterica]
dedRef_860782-860783_23	4	Short	6	1,50845E-22	WP_205903887	hypothetical protein, partial [Pseudomonas viridiflava]
dedRef_917560-917560_23	3	Short	5	2,57572E-21	WP_213055896	hypothetical protein, partial [Escherichia coli]
dedRef_199573-199573_55	6	Short	7	8,06871E-21	XP_063455816	uncharacterized protein LOC134729584 [Pan paniscus]
dedRef_724392-724392_21	7	Tall	6	1,6404E-19	WP_207838564	hypothetical protein, partial [Pseudomonas sp. 43(2021)]
dedRef_1479655-1479656_31-2	1	Short	5	2,6909E-19	WP_212578393	hypothetical protein, partial [Vibrio parahaemolyticus]
dedRef_1846774-1846774_54	14	Tall	65	8,59651E-18	WP_214645256	hypothetical protein, partial [Mycobacterium tuberculosis]
			Read 1 - CHG			
dedRef_732447-732447_37	6	Tall	5	9,21577E-26	WP_227645212	hypothetical protein, partial [Klebsiella pneumoniae]
dedRef_377192-377193_42-1	11	Short	5	8,37072E-24	WP_227645212	hypothetical protein, partial [Klebsiella pneumoniae]
dedRef_689981-689982_35	1	Short	26	1,58498E-23	WP_323510696	hypothetical protein, partial [Cryobacterium sp. 10S3]

dedRef_433436-433437_25	7	Short	5	2,38964E-23	WP_217884343	DUF1725 domain-containing protein [Thioflexithrix psekupsensis]
dedRef_2496177-2496179_44-1	15	Tall	5	2,64361E-23	XP_036016766	igE-binding protein-like [Mus musculus]
dedRef_1872316-1872317_3-2	22	Short	5	3,03145E-23	WP_311050758	reverse transcriptase domain-containing protein [Enterococcus faecalis]
dedRef_1478089-1478092_23	12	Short	5	5,41014E-22	XP_006498466	rho GTPase-activating protein 15 isoform X5 [Mus musculus]
dedRef_473265-473265_34	7	Short	5	7,51216E-22	WP_334412441	hypothetical protein, partial [Enterococcus faecium]
dedRef_695459-695459_23-1	4	Short	5	1,04402E-21	WP_334412441	hypothetical protein, partial [Enterococcus faecium]
dedRef_1314859-1314859_34	5	Short	5	2,23411E-21	WP_228003977	hypothetical protein, partial [Leptospira borgpetersenii]
				Read 2 - CHG		
dedRef_179953-179954_18	1	Short	8	2,67984E-20	WP_206338328	hypothetical protein, partial [Pseudomonas viridiflava]
dedRef_18060-18060_20	4	Short	5	4,5495E-20	YP_003541050	ATP synthase F0 subunit 6 [Homo sapiens subsp. 'Denisova']
dedRef_3069618-3069619_3	2	Tall	5	2,97206E-12	WP_287256559	DDE-type integrase/transposase/recombinase, partial [Stenotrophomonas sp.]
dedRef_4956546-4956546_20	2	Tall	43	3,03115E-11	WP_203744308	hypothetical protein, partial [Escherichia coli]
dedRef_1010182-1010183_35	2	Short	5	3,89727E-10	YP_008082327	RNA polymerase beta' subunit [Pinus taeda] >ref|YP_010449301.1| RNA polymerase beta' subunit [Pinus rigida]
dedRef_973251-973252_36	3	Short	5	9,72089E-10	XP_020262272	uncharacterized protein LOC109838225 [Asparagus officinalis]
dedRef_1369164-1369165_24	3	Short	5	2,82479E-08	WP_254062609	hypothetical protein [Enterococcus faecalis]
dedRef_80571-80572_33	3	Short	5	3,47192E-08	WP_294211963	hypothetical protein, partial [uncultured Sphingomonas sp.]
dedRef_1509534-1509535_29-1	3	Short	5	1,48577E-07	WP_267992283	reverse transcriptase family protein, partial [Actinobacillus pleuropneumoniae]
dedRef_582116-582117_24	3	Short	5	1,51422E-07	WP_214280456	hypothetical protein [Escherichia coli]
				Read 1 - CHH		
dedRef_2480-2482_31-2	49	Short	5	7,26458E-28	WP_311050758	reverse transcriptase domain-containing protein [Enterococcus faecalis]
dedRef_383569-383570_37	49	Short	5	1,46277E-27	WP_330949972	reverse transcriptase domain-containing protein [Streptococcus anginosus]

dedRef_3817988-3817990_34-2	44	Short	5	1,57627E-23	WP_330949811	endonuclease/exonuclease/phosphatase family protein [Streptococcus anginosus]
dedRef_689981-689982_35	38	Short	26	1,58498E-23	WP_323510696	hypothetical protein, partial [Cryobacterium sp. 10S3]
dedRef_2496177-2496179_44-1	42	Tall	5	2,64361E-23	XP_036016766	igE-binding protein-like [Mus musculus]
dedRef_1649087-1649089_45	22	Tall	5	1,02057E-22	NP_775626	pseudouridylate synthase RPUSD2 [Mus musculus]
dedRef_1112190-1112190_24-2	50	Short	5	1,04308E-22	WP_334412441	hypothetical protein, partial [Enterococcus faecium]
dedRef_32467-32467_31	37	Short	5	7,28239E-22	WP_227645862	hypothetical protein, partial [Klebsiella pneumoniae]
dedRef_695459-695459_23-1	42	Short	5	1,04402E-21	WP_334412441	hypothetical protein, partial [Enterococcus faecium]
dedRef_22135-22137_26	42	Short	5	7,4278E-21	WP_228029457	reverse transcriptase domain-containing protein, partial [Leptospira borgpetersenii]
				Read 2 - CHH		
dedRef_3510320-3510321_2	5	Tall	5	1,75185E-13	WP_271080373	proton-conducting transporter membrane subunit, partial [Enterobacter hormaechei]
dedRef_802343-802344_32	27	Tall	5	1,02822E-08	WP_267992538	reverse transcriptase domain-containing protein, partial [Actinobacillus pleuropneumoniae]
dedRef_601720-601721_35	24	Tall	5	2,00566E-07	WP_214280456	hypothetical protein [Escherichia coli]
dedRef_633310-633311_31-16	20	Short	5	7,03037E-07	WP_214280456	hypothetical protein [Escherichia coli]
dedRef_812402-812403_29	17	Tall	5	1,14106E-06	WP_254602374	NAD-dependent succinate-semialdehyde dehydrogenase, partial [Burkholderia lata]
dedRef_613800-613801_27	13	Tall	5	1,44263E-06	WP_158405218	SMC-Scp complex subunit ScpB, partial [Faecalibacterium prausnitzii]
dedRef_1039075-1039076_27	15	Tall	5	1,63478E-06	WP_254062609	hypothetical protein [Enterococcus faecalis]
dedRef_1211072-1211072_23	16	Tall	5	2,4726E-06	WP_254602374	NAD-dependent succinate-semialdehyde dehydrogenase, partial [Burkholderia lata]
dedRef_723056-723057_22	11	Tall	5	2,91862E-06	WP_214280456	hypothetical protein [Escherichia coli]
dedRef_1163173-1163174_27	13	Tall	5	2,98082E-06	WP_158405218	SMC-Scp complex subunit ScpB, partial [Faecalibacterium prausnitzii]

Table S17. BLASTx of TvS showing top 10 hits of differentially methylated motifs for read 1 and read 2 together.

meth.ID	Query	Number of HSPs	Lowest E-value	Accession	Description (E-value)
		Read 1 - CpG			
345947/345949/345951	dedRef_2497661-2497662_35	5	1,39857E-28	WP_098733516	replication protein P, partial [Escherichia coli]
1564181/1564179	dedRef_857302-857303_35	5	2,03381E-28	WP_252146668	excisionase [Escherichia coli]
1514083	dedRef_714684-714685_43	5	1,02144E-27	WP_032254233	replication protein P, partial [Escherichia coli]
638591	dedRef_136580-136581_35	5	3,11153E-27	WP_187267031	terminase gpA endonuclease subunit, partial [Escherichia coli]
236122	dedRef_389796-389796_24-2	6	1,67224E-24	WP_165766559	hypothetical protein, partial [Staphylococcus capitis]
401548	dedRef_1430190-1430191_24	37	4,31959E-23	WP_331708159	hypothetical protein, partial [Pseudomonas aeruginosa]
714550	dedRef_1649087-1649089_45	5	1,02828E-22	NP_775626	pseudouridylate synthase RPUSD2 [Mus musculus]
1423149	dedRef_4941418-4941418_16	11	1,14991E-21	WP_021666093	MULTISPECIES: hypothetical protein [Bacteria]
1591147	dedRef_948692-948693_56	5	2,71438E-20	NP_008211	cytochrome b [Pan paniscus]
265568	dedRef_1613544-1613544_33	5	4,21209E-20	WP_228003977	hypothetical protein, partial [Leptospira borgpetersenii]
		Read 2 - CpG			
34443	dedRef_4349474-4349475_27-6	5	3,3289E-26	XP_055219682	Bardet-Biedl syndrome 4 protein isoform X5 [Gorilla gorilla gorilla]
140684	dedRef_3501534-3501534_25	5	4,16852E-25	WP_146136603	RBD-like domain-containing protein [Aphanothece hegewaldii]
763502	dedRef_25086-25087_42	5	5,45474E-25	XP_036010831	contactin-5 isoform X1 [Mus musculus]
31033	dedRef_932878-932878_38-1	5	1,04694E-24	XP_054385603	DNA mismatch repair protein Mlh3 isoform X7 [Pongo abelii]
28922	dedRef_1587864-1587864_34-7	5	2,91491E-24	XP_042750896	protein GVQW1-like, partial [Lagopus leucura]
339202	dedRef_592472-592473_24-2	5	3,54782E-24	WP_323567329	hypothetical protein, partial [Pseudomonas sp. CCI2.4]
79766/79764	dedRef_369428-369430_45	5	1,18358E-23	XP_054961064	protein FRG1B-like [Pan paniscus]
10587	dedRef_2731331-2731332_7-3	6	4,10855E-23	WP_206671226	hypothetical protein, partial [Staphylococcus aureus]
15889	dedRef_1597560-1597560_32-1	5	2,79757E-22	XP_054385603	DNA mismatch repair protein Mlh3 isoform X7 [Pongo abelii]

257392	dedRef_848433-848433_51	5	1,16941E-21	XP_054961064	protein FRG1B-like [Pan paniscus]
Read 1 - CHG					
2163495	dedRef_857302-857303_35	5	2,03381E-28	WP_252146668	excisionase [Escherichia coli]
268805/268807	dedRef_498143-498144_54	5	3,23499E-28	WP_174714868	C40 family peptidase, partial [Escherichia coli]
524317	dedRef_2013374-2013382_30	5	5,31202E-28	WP_146627548	antitermination protein, partial [Streptococcus pneumoniae]
2045831/2045833/2045835/2045821/2045829/2045826	dedRef_714684-714685_43	5	1,02144E-27	WP_032254233	replication protein P, partial [Escherichia coli]
835373	dedRef_13123-13124_50	5	6,90857E-27	WP_136694900	tail assembly protein, partial [Escherichia coli]
62331	dedRef_1587864-1587864_34-7	5	2,93692E-24	XP_042750896	protein GVQW1-like, partial [Lagopus leucura]
1717804	dedRef_4285559-4285560_31	5	1,18508E-23	WP_240491102	MULTISPECIES: hypothetical protein [Bacilli]
430179	dedRef_2797613-2797614_23	5	1,34138E-23	WP_218641054	endonuclease/exonuclease/phosphatase family protein, partial [Paenibacillus odorifer]
261298	dedRef_2496177-2496179_44-1	5	2,66357E-23	XP_036016766	igE-binding protein-like [Mus musculus]
880719	dedRef_1418514-1418514_40	5	8,35351E-22	XP_045002940	LOW QUALITY PROTEIN: uncharacterized protein LOC123460056, partial [Jaculus jaculus]
Read 2 - CHG					
173906	dedRef_530079-530080_29	5	1,10171E-25	WP_143768019	hypothetical protein, partial [Paenibacillus odorifer]
298355	dedRef_831942-831942_23	5	4,96532E-24	WP_139312580	hypothetical protein, partial [Pseudomonas aeruginosa]
191754	dedRef_720424-720425_20-3	5	8,43125E-24	WP_223009790	hypothetical protein, partial [Stenotrophomonas maltophilia]
181269	dedRef_1584780-1584781_38	5	1,05765E-23	WP_219861017	hypothetical protein, partial [Corynebacterium diphtheriae]
133759	dedRef_699133-699133_35-1	5	8,42362E-23	WP_000352757	reverse transcriptase family protein, partial [Helicobacter pylori]
306529	dedRef_1060000-1060000_33-2	5	1,0509E-22	WP_206603989	hypothetical protein, partial [Vibrio vulnificus]
19040	dedRef_1597560-1597560_32-1	5	2,79757E-22	XP_054385603	DNA mismatch repair protein Mlh3 isoform X7 [Pongo abelii]
1849144	dedRef_4229842-4229843_31-2	5	3,28846E-22	WP_205976132	DUF1725 domain-containing protein, partial [Pseudomonas viridiflava]

241514	dedRef_4972335-4972337_27-1	5	4,11348E-22	WP_258304901	hypothetical protein, partial [Escherichia coli]
241703	dedRef_5004293-5004293_30	5	9,52036E-22	WP_139372712	hypothetical protein, partial [Bacillus anthracis]
			Read 1 - CHH		
458959	dedRef_375863-375864_24	5	4,69301E-25	WP_281302239	hypothetical protein, partial [Iodidimonas sp. MBR-14]
2619715	dedRef_2268147-2268148_35	6	5,10597E-25	WP_141607967	hypothetical protein [Leptospira interrogans]
655326	dedRef_1059269-1059270_27-2	7	4,49854E-24	WP_219553321	hypothetical protein, partial [Klebsiella pneumoniae]
839254/839294	dedRef_899225-899226_33	5	7,0827E-24	XP_047953659	30S ribosomal protein S12-B, chloroplastic [Salvia hispanica]
1964729	dedRef_1430787-1430787_21	37	1,61182E-23	WP_157464686	hypothetical protein [Crocosphaera chwakensis]
498649	dedRef_780728-780729_49-2	23	2,86151E-23	WP_213842745	hypothetical protein, partial [Enterobacter hormaechei]
1075659	dedRef_1434088-1434089_22	5	2,95952E-23	WP_303704525	hypothetical protein, partial [Klebsiella pneumoniae]
1269561	dedRef_4150744-4150745_22	5	6,99859E-23	WP_142951599	hypothetical protein, partial [Bacillus thuringiensis]
494862	dedRef_552007-552011_32-1	5	9,05656E-23	WP_206682097	DUF1725 domain-containing protein, partial [Escherichia coli]
637207	dedRef_494841-494842_29-2	5	9,5152E-23	WP_142319120	hypothetical protein, partial [Bacillus cereus]
			Read 2 - CHH		
1463819/1463834	dedRef_720879-720880_33	5	7,07049E-27	WP_227645212	hypothetical protein, partial [Klebsiella pneumoniae]
873381	dedRef_590435-590437_32	5	4,81022E-26	WP_218140791	hypothetical protein [Cylindrospermopsis raciborskii]
6958622	dedRef_1649422-1649422_34	5	7,94701E-26	XP_036016762	igE-binding protein-like [Mus musculus]
443754	dedRef_394207-394212_34-3	5	1,96797E-24	XP_054112079	ubiquitin carboxyl-terminal hydrolase 32 isoform X2 [Callithrix jacchus]
2477134	dedRef_3186886-3186887_14	28	1,12757E-21	WP_232783055	hypothetical protein, partial [Bacillus sp. BA3]
216018	dedRef_952293-952293_43-1	5	4,05735E-21	XP_035169934	protein GVQW1-like, partial [Oxyura jamaicensis]
1085615	dedRef_801010-801011_53	78	5,7303E-21	WP_305212016	hypothetical protein, partial [Klebsiella variicola]
13044749	dedRef_865645-865647_53	29	6,03534E-21	WP_141394363	hypothetical protein [Enterobacter hormaechei]

10679562	dedRef_4864839-4864839_13	5	6,97707E-18	WP_208399076	hypothetical protein, partial [Staphylococcus epidermidis]
4971534	dedRef_1334571-1334572_30	5	1,08936E-16	WP_206682097	DUF1725 domain-containing protein, partial [Escherichia coli]