International Journal of Innovative Research and Advanced Studies (IJIRAS)
Volume 7 Issue 5, May 2020

ISSN: 2394-4404

# Review On Status Of Association Studies In Forestry

**Ayushman Malakar**

M.Sc. Forestry, Forest Research Institute, Dehradun, Uttarakhand

***Abstract:*** ***Forests are the source of raw material for mitigating several essential needs of humans, including building materials, paper products, firewood for heat and cooking and many tree-crop foods. Improvement of forestry crops are hence a dire need of modern silvicultural practices. Genomic research on forest trees is motivated by the need to support genetic improvement programmes and develop diagnostic tools for the conservation, restoration and management of natural populations. But long regeneration times and large genome structures, the lack of well characterized mutations for reverse genetic approaches, and lack of proper model species hinders the advanced genetic researches in forest trees. Conventional linkage mapping has not proved to be efficient in tree improvements. With advent of next-generation sequencing technologies, researches have shown an enormous interest in using association mapping to identify genes responsible for quantitative variation of complex traits in forest trees. In comparison of conventional marker assisted breeding methodologies, Association mapping takes the advantage of Linkage Disequilibrium (LD)- the non-random associations of loci in haplotypes, by which the complex traits of forest trees can be exploited for tree improvement through association studies. Moreover, association studies also help to avoid the false detection rates during complex trait dissection. It can be predicted that association studies in forest trees will open new horizons for researches in near future and tree breeding technologies will reach another breakthrough as tree improvement of several economically important tree species will be very much feasible.***

***Keywords: association mapping; candidate genes; linkage disequilibrium; forest trees; tree improvement***

## I. INTRODUCTION

Forest trees constitute about 82% of the continental biomass and harbour more than 50% of the terrestrial biodiversity (Roy, Saugier and Mooney, 2001). Forests are the source of raw material for mitigating several essential needs of humans, including building materials, paper products, firewood for heat and cooking and many tree-crop foods. Forest trees also provide various ecological services such as preservation of biodiversity, carbon sink, climate regulation and preservation of water quality and represent our cultural and patrimonial heritage (UNEP. Vital Forest Graphics, 2009). Genomic research on forest trees is motivated by the need to support genetic improvement programmes and develop diagnostic tools for the conservation, restoration and management of natural populations (Neale and Kremer, 2011). Genetic researches in forest trees has been hindered by their long regeneration times and large genome structures, the lack of well characterized mutations for reverse genetic approaches, and limited funding. Selection of trees with superior traits of economically importance for plantation and breeding programs is a very lengthy and expensive process for the tree breeders, when the reproductive cycle is long and selection is based on physical traits.

However, the researchers have achieved great success in the improvement of tree species with the advent of modern molecular genomics. Marker Assisted Selection (MAS) has been proven to have a great potential in providing rapid and effective selection. But absence of full-sib progenies is the greatest limitation in linkage mapping of important traits in forest trees. In such situations, Association Mapping (AM), based on associations between genotype and phenotype variation in unorganized natural populations, proves to be the savior. AM takes advantage of both Linkage Disequilibrium (LD) and historical recombination present within the gene pool of an organism, utilizing a broader reference population (Ersoz and Buckler, 2007; Myles et al., 2009) and thus genotypic and phenotypic correlations are investigated in

unrelated individuals (Khan and Korban, 2012). LD refers to a historically reduced (nonequilibrium) level of the recombination of specific alleles at different loci controlling particular genetic variations in a population which can be detected statistically, and has been widely applied to map and eventually clone a number of genes underlying the complex genetic traits (Abdurakhmonov and Abdukarimov, 2008). Advantages of this approach over the biparental approach are that much larger and more representative gene pools may be surveyed without any need of developing biparental mapping populations, with higher resolution (Remington et al., 2001). AM is particularly suited for forest trees and perennial horticultural crops to overcome their characteristic pedigree-based mapping limitations (Khan and Korban, 2012). Conifers like pine and spruce species, poplars and eucalypts are some of the forest trees which have been studied extensively under LD-based Association Mapping for genotypic improvement by several researchers all over the world.

## II. ASSOCIATION MAPPING AS AN ALTERNATIVE APPROACH IN TREE GENOMICS

In trees, a few notable examples of single-gene (that is, mendelian) traits are known, including the identification of the major gene for resistance to white-pine blister rust (*Cronartium ribicola*), which is found in many species of the subgenus *strobus* of the genus *Pinus* (Kinloch et al., 1970). However, the vast majority of traits of both economic and ecological interest in forest trees are complex, quantitative traits. These include growth and yield, the properties of wood, resistance to diseases and insects, and resistance or tolerance to abiotic stresses (White et al., 2007). Forest geneticists have been studying the heritability of these traits for several years, there are numerous literatures regarding these studies. In recent years, forest geneticists have moved to an alternative approach for complex trait dissection: association genetics, as this approach has several advantages in many forest tree species, principally the lack of significant population structure in association genetics populations and the rapid decay of linkage disequilibrium (Neale and Savolainen, 2004).

Association Mapping (AM) serves as a viable alternative approach that can overcome the limitations of pedigree-based mapping in perennial plants. Turning the gene-tagging efforts from biparental crosses to natural population of lines (or germplasm collections), and from traditional QTL mapping to linkage disequilibrium (LD)-based association study became a powerful tool in mapping of the genes of interest (Goldstein and Weale, 2001). The overall approach of population-based association mapping in plants varies based on the methodology chosen, assuming structured population samples, the performance of association mapping includes the following steps as depicted by Abdurakhmonov and Abdukarimov, 2008 (Figure 1) as; (1) Selection of a group of individuals from a natural population or germplasm collection with wide coverage of genetic diversity; (2) Phenotyping i.e. recording or measuring the phenotypic characteristics (yield, quality, tolerance, or resistance) of selected population groups, preferably, in different environments and multiple replication or trial design; (3) Genotyping a mapping population individuals with available molecular markers; (4) Quantification of the extent of LD of a chosen population genome using a molecular marker data; (5) Assessment of the population structure (the level of genetic differentiation among groups within a sampled population individuals) and kinship (coefficient of relatedness between pairs of each individuals within a sample); and (6) Based on information gained through quantification of LD and population structure, correlation of phenotypic and genotypic or haplotypic data with the application of an appropriate statistical approach that reveals "marker tags" positioned within close proximity of targeted trait of interest. Consequently, a specific gene(s) controlling a QTL of interest can be cloned using the marker tags and annotated for an exact biological function. As a starting point for association mapping, it is important to gain knowledge of the patterns of LD for genomic regions of the "target" organisms and the specificity of the extent of LD among different populations or groups to design and conduct unbiased association mapping (Nordborg et al., 2002).

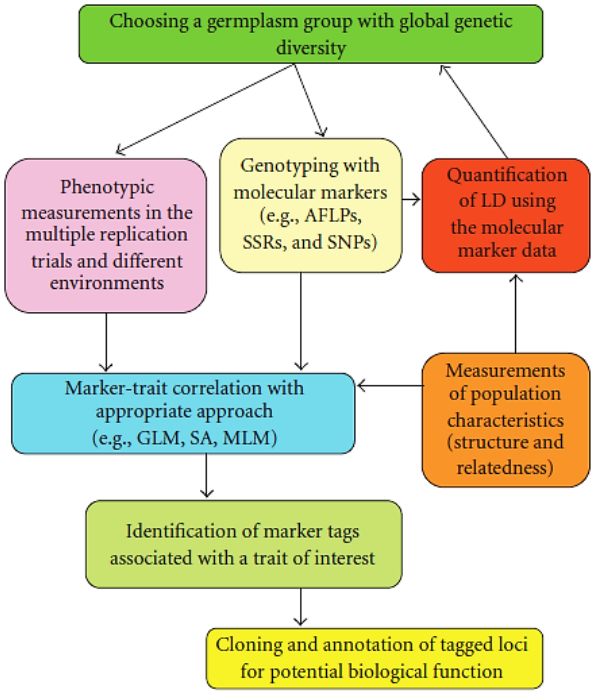


*Figure 1: The scheme of association mapping for tagging a gene of interest using germplasm accessions (After, Abdurakhmonov and Abdukarimov, 2008)*

## III. DESIGNS AND MODELS FOR ASSOCIATION MAPPING

Given the geographical origins, local adaptation, and breeding history of assembled genotypes in an association mapping panel, the non-independent samples usually contain both population structure and familial relatedness (Yu and Buckler, 2006). Recently, several statistical methods have been proposed to account for population structure and familial relatedness, structured association (SA) (Falush et al., 2003), genomic control (GC) (Devlin and Roeder, 1999), mixed model approach (Yu et al., 2006), and principal component

approach (Price et al., 2006). With these methods, the issue of false positives generated by population structure can now be dealt with accordingly (Yu et al., 2006). AM can be performed in all panmictic populations that harbour considerable LD in genomic regions involved in the control of the target phenotypic traits. The Nested Association Mapping (NAM) population can be used for both linkage mapping of QTLs and AM as it uses RILs developed from a diverse set of parents, requires a smaller number of markers than GWAS in population-based association panels, and has higher resolution than QTL linkage mapping. (Yu et al., 2008). The Multiparent Advanced Generation Intercross (MAGIC) populations comprise a set of RILs derived from a complex cross or a set of crosses or several recombination involving multiple parents which can be used for both linkage and association mapping of multiple traits for which the parents differ, and multiple alleles at the target loci can also be detected (Descalsota et al., 2018).

Initially the AM strategies were drawn for human genetic studies and these were applied to plant genetic studies without much modification. Subsequently, more precise and powerful methods for unbiased AM in plants were developed. There are several different approaches for the detection of significant LD, ranging from the simple chi-square test through analysis of variance to complex likelihood-based procedures (Singh and Singh, 2015). When the LD between a marker and a QTL is strong, the various methods would give comparable results. AM for quantitative traits in plants is based on regression, maximum likelihood, and Bayesian approaches (Oraguzie et al., 2007). There are several experimental designs used depending upon from the models test for association between a single-marker locus and a single target trait at a time to models for simultaneous evaluation of multiple marker loci as well as multiple traits shown in the Table 1.

| Design | Features | Remarks |
|---|---|---|
| **Case and control approach** | Based on a group carrying the disease causing allele (cases) and an unrelated group of equal size lacking the disease (control) | Used in humans; modifications like HRR, genomic control |
| **Transmission disequilibrium test** | A family-based design; compares transmission versus nontransmission of the marker to the affected progeny from one heterozygous and one homozygous parent | Used in humans |
| **Structured association** | Designed to minimize the effects of population structure; one version is the general linear model (GLM) | GLM implemented in TASSEL |
| **Mixed linear model (MLM)** | Designed to minimize the effects of population structure and kinship; markers and *Q* treated as fixed effects, while background QTLs are treated as random effects | Uses *K* or both *Q* and *K* matrices; EMMA is an improved version of mixed model |
| **Multilocus mixed model (MLMM)** | Multiple loci used as cofactors in the model; uses stepwise mixed model regression for the selection of loci and an approximate version of mixed model of correction for population structure | More QTL detection power and lower FDR than single locus tests |
| **Multitrait mixed model (MTMM)** | Simultaneous analysis of two or more correlated traits using the mixed model; separates genetic and environmental correlations and corrects for population structure | More power than single trait models when the traits are correlated; otherwise, lower power |
| **Joint linkage-association mapping** | Analysis of a sample drawn from a natural population and the open-pollinated progeny from this sample | Uses both LD and linkage analysis |
| **Nested association mapping (NAM)** | LD and linkage mapping in NAM populations | Higher power than AM alone |

*Table 1: List of experimental designs used for association mapping (After, Singh and Singh, 2015)*

## IV. LINKAGE DISEQUILIBRIUM

The concept of LD was first described by Jennings in 1917, and its quantification (D) was developed by Lewtoninin (1964). Linkage disequilibrium is also referred as "Gametic Phase Disequilibrium" (GPD) or "Gametic Disequilibrium" (GLD) that describes the nonrandom association of haplotypes within unrelated populations with a distantly shared ancestry, assuming Hardy-Weinberg equilibrium (HWE) (Abdurakhmonov and Abdukarimov, 2008). The simplified explanation of the commonly used LD measure, D or $\acute{D}$ (standardized version of D), is the difference between the observed gametic frequencies of haplotypes and the expected gametic haplotype frequencies under linkage equilibrium ($D = P_{AB} - P_A P_B = P_{AB}P_{ab} - P_{Ab}P_{aB}$) (Oraguzie et al., 2007).

Disequilibrium is the result of disturbing effects of one or more of the evolutionary factors on gene and genotype frequencies in the population. Disequilibrium can also result from linkage between the genes *a* and *b*. The term *linkage disequilibrium* (LD) signifies that a specific allele at one locus occurs with a specific allele at the second locus more often than expected on the basis of random assortment of the two loci. The two loci may represent two markers, two genes/QTLs, or one gene/QTL and one marker. Thus, in simple terms, LD describes a non-random association between alleles of two or more loci. As a result, the allelic combinations of the concerned loci observed in the population deviate significantly from their frequencies expected on the basis of independent assortment. In each generation of random mating, the magnitude of "*d*" will decline by the value "*rd*", where *r* is the frequency of recombination between the two loci. This decline in LD is known as LD decay. Further, the magnitude of LD will decrease with the genetic distance

between the two loci since it is inversely related to the frequency of recombination between them. In each generation, there will be recombination between the two loci during meiosis, which will lead to a decline in the magnitude of LD. In simple terms, LD between two loci decays both temporally (as the generation advances) and spatially (with the increasing distance between the two loci) (Singh and Singh, 2015, p. 227).

Choosing the appropriate LD measures really depends on the objective of the study, and one performs better than other in particular situations and cases; however, $\acute{D}$ and $r^2$ is the most commonly used measures of LD (Oraguzie et al., 2007). The statistical significance of LD estimates is determined by Fisher's exact test when the two loci have two alleles each and by multifactorial permutation analysis when more than two alleles occur at one or both the loci (Flint-Garcia et al., 2003). The merits, sensitivity, comparison, appropriate statistical tests, and calculation methodology for these LD measures with the utilization of bi-allelic or multi-allelic loci have been extensively reviewed by Gupta et al. (2005). In association mapping, efforts are made to filter out all other influences on LD estimates to, ideally, retain the effects of only linkage and use this information for identification of markers closely linked to the genes/QTLs governing the trait(s) of interest (Singh and Singh, 2015). The probability of detecting "true" marker-trait associations in a sample using AM is called power of association mapping (Singh and Singh, 2015). The power of an AM experiment depends on several factors, including the extent and evolution of LD in the population, nature of gene effects involved in control of the target trait, sample size, experimental design, accuracy of phenotyping, type of markers, etc. (Churchill and Doerge, 1994). The chances of detecting LD are the greatest for mutations that are of recent origin (i.e., are in strong LD), have large effect on the phenotype, and are present in a relatively less frequent haplotype background (Singh and Singh, 2015). The power of LD detection can be markedly increased by choosing a suitable study design, reducing the environmental variation, and increasing the genetic effects by selecting the extreme phenotypes (Ball, 2005). The marker genotyping work can be reduced by using such genomic regions for mapping that have known QTLs/candidate genes, selecting one marker from each haplotype of interest, etc. (Singh and Singh, 2015). A large sample size would be required to enhance the reliability of associations and required sample size for a given power can be estimated on the basis of Bayes factors using R function of *ld.design* from *ldDesign package* (Ball, 2007).

## V. APPROACHES IN ASSOCIATION MAPPING

Based on the scale and focus of a particular study, association mapping generally falls into two broad categories, (i) Genome-Wide Association Mapping, or genome scan, which surveys genetic variation in the whole genome to find signals of association for various complex traits; and (ii) Candidate-Gene Association Mapping, which relates polymorphisms in selected candidate genes that have purported roles in controlling phenotypic variation for specific traits (Risch and Merikangas, 1996).

In Genome-Wide Association Studies (GWAS), the markers used for genotyping are distributed, preferably evenly and densely, over the whole genome (Singh and Singh, 2015). In this approach, all the loci involved in the control of all the traits showing variation in the sample can be evaluated at once. It is important that a genome-wide linkage map of markers of the concerned species must be available to permit the selection of an appropriate set of markers. Accurate phenotyping is a prerequisite for GWAS for arriving at valid conclusions as an increase in the number of individuals or lines included for phenotyping enhances the power of AM much more than an increase in the number of markers used for genotyping (Ingvarsson and Street, 2011).

Whereas, Candidate-Gene Association Mapping is a hypothesis driven approach to complex trait dissection, with biologically relevant candidates selected and ranked based on the evaluation of available results from genetic, biochemical, or physiology studies in model and non-model plant species (Mackay, 2001; Risch and Merikangas, 1996). In candidate gene approach, marker analysis is restricted to the genomic regions having the candidate genes. A candidate gene is a gene that is expected, on the basis of previous knowledge, to be involved in the control of a trait of interest (Singh and Singh, 2015). The candidate genes are identified on the basis of the information analyzed from different sources like comparative genomics, genome sequence annotation, transcript profiling, QTL analysis, etc. Instead of phenotyping, in candidate gene approach, genotyping is focused in the genomic regions with the candidate genes which greatly reduces the target genomic region which can be analyzed with a high density of markers as the total number of markers used as well as the sample size will also be considerably reduced (Stich et al., 2008). Candidate gene approach has been used to identify genes involved in the control of many traits, including morphological, phenological, and stress resistance traits (Ingvarsson and Street, 2011). This approach may be able to identify a QTL where genome-wide AM fails to detect a significant marker-trait association after false discovery rate (FDR) correction is applied (Singh and Singh, 2015). In addition, the use of this approach along with GWAS tends to increase the power and precision of QTL detection (Gupta et al., 2014).

## VI. CURRENT STATUS OF ASSOCIATION STUDIES IN FOREST TREES

The pioneer association studies in plants were performed by Beer et al. (1997) in oat, and by Virk et al. (1996) in rice. Although there is wide variation among different forest trees, generally they are out-crossing, long-lived, and at early stages of domestication (Savolainen and Pyhajarvi, 2007). Despite similarities among forest trees, there are large differences between genetic and lifecycle characteristics of forest trees that render some better suited for genetic studies than others (Khan and Korban, 2012). Availability of genomic resources is one of the key limitations for conducting an AM study in a crop. Genome sizes of forest trees vary considerably, impacting availability of genomic resources, and influencing AM strategies (Tuskan et al., 2006). LD declines rapidly in

forest trees, within 1 kb, compared to that of self-pollinated plant species, thus requiring availability of large numbers of molecular markers for AM. In addition to the genome sequence of poplar, large numbers of expressed sequence tag (EST) sequences are also available for multiple forest trees (Khan and Korban, 2012). So far, candidate gene-based AM studies and population genetic neutrality tests have frequently been used to study wood-related economic and adaptive traits, and to investigate gene behaviour under natural selection conditions in forest trees (Neale, 2007; Eckert et al., 2009a). Probably the first notable association studies in forest trees is done by Thumma et al. (2005) where associations of polymorphisms in *Cinnamoil CoA Reductase (CCR)* with early wood micro fibril angle trait and polymorphisms a putative stress response gene with wood density and wood growth rate were reported.

## ✓ ASSOCIATION STUDIES IN CONIFERS

SNPs are generally identified by resequencing ESTs using a small SNP discovery panel and are then genotyped on large samples of unrelated trees by high-throughput genotyping assays (Pavy et al., 2008; Eckert et al., 2009b). Sufficient numbers of EST sequences are available for *Picea* and *Pinus* (Pavy et al., 2005; Rungis et al., 2005) to pursue SNP discovery and conduct LD studies in conifers.

Gonza´lez- Mart´ınez et al. (2007) conducted the first multi-gene association study in forest trees. A population of 422–435 unrelated loblolly pine trees (*P. taeda*) trees, in a clonally replicated trial, was used to conduct an association analysis study of 58 SNPs, from 20 wood- and drought-related candidate genes and wood property traits. These traits included earlywood and latewood specific gravity, percentage latewood, earlywood microfibril angle, and wood chemistry (lignin and cellulose contents). They used mixed linear models (MLM) to perform AM analysis, where population structure and relatedness was accounted for. It was suggested that due to rapid LD decay in conifers, SNPs revealing genetic associations were likely to be located in close proximity to causative polymorphisms (Gonza´lez- Mart´ınez et al., 2007). A strong association between a SNP within the candidate gene *4-coumarate CoA ligase (4cl)* and percentage latewood was also detected in this study, thus confirming previous findings based on co- location of a QTL for percentage latewood. In another study, several candidate gene associations were detected together with two very promising associations, a cell structure stabilizing *dehydrin gene (dhn-1)* and a cell wall reinforcement protein gene (*lp5*) (Gonza´lez-Mart´ınez et al., 2008). This study demonstrated successful candidate gene association mapping in trees using a complex family structure.

Pitch canker, a disease caused by the necrotrophic pathogen *Fusarium circinatum*, is an important fungal disease of loblolly pine. Quesada et al. (2010) have identified genes that are associated with resistance to pitch canker in loblolly pine using a set of 498 largely unrelated clonally propagated genotypes. Significant associations have been re- examined in an Australian land race, and those associations previously observed in the discovery population have further validated associations of two genes with wood density. Decreased wood density is associated with a minor allele, thus suggesting that these SNPs may be under weak negative purifying selection for wood density. These findings clearly demonstrate the utility of LD mapping in detecting associations, even when the power of detecting SNPs with small effects is anticipated to be low (Dillon et al., 2010).

Heuertz et al. (2006) have conducted LD and tests of neutrality in spruce (*Picea*) and have surveyed DNA polymorphisms at 22 loci in; 47 haplotypes from seven populations. Their results have revealed that the overall nucleotide variation is limited, being lower than that observed in most plant species. LD is also restricted and does not extend beyond a few hundred base pairs. Recently, neutrality tests have been used to study the effects of natural selection on 41 candidate genes from loblolly pine; these genes have been selected primarily from host– pathogen interactions together with 15 drought tolerance and 13 wood-quality genes identified in previous studies (Ersoz et al., 2010). Dillon et al. (2010) have evaluated the utility of LD mapping to detect associations between SNPs and wood quality in a natural population of *Pinus radiata*. To further elucidate the genetic control of adaptive traits in Douglas fir (*Pseudotsuga menziesii*), Krutovsky and Neale (2005) studied LD and haplotype and nucleotide frequencies, and performed neutrality tests in cold hardiness and wood quality-related candidate genes.

Recently, a total of 240 genotypes of *Pinus roxburghii* (Himalayan Chir Pine) from a natural population in Chakrata division (Tiunee range), Uttarakhand (India) were evaluated for resin yield (Rawat et al., 2014). 53 genotypes were selected after excluding the individuals with similar resin production. The selected 53 individuals were best representatives of the variation in resin yield in Chakrata population which varied between 0.25 and 8.0 kg/tree/year and were used for genotyping and association analysis using SSR markers. A total of 19 polymorphic SSRs (11 cpSSR and 8 nSSR) were used in the study.

## ✓ ASSOCIATION STUDIES IN POPLAR

Till Date, poplar (*Populus trichocarpa*) is the only forest tree with a complete genome sequence, thus allowing resequencing of different genotypes to identify SNPs for genetic studies. Ingvarsson et al. (2006) first reported on genetic dissection of complex adaptive traits using candidate genes in poplar. They identified SNPs in the phytochrome gene (*phyB2*) that co-located to a previously reported QTL for timing of bud set, and this was genotyped in 16 *Populus tremula* populations collected along a latitudinal gradient in Sweden.

Hall et al. (2007) also found several phenological traits of *P. tremula* with strong genetic differentiation and clinal variation across the latitudinal gradient. Ingvarsson et al. (2008) reported that polymorphism varied substantially across the *phyB2* region and proposed that due to low LD in this region, these SNPs were strong candidates that were causally linked to variation in bud set. Porth et al. (2013) performed GWAS for key wood chemistry and ultrastructure traits in a population of 334 unrelated *Populus trichocarpa.*

✓ ASSOCIATION STUDIES IN EUCALYPTUS

Eucalyptus (*Eucalyptus nitens*) is a widely adapted forest tree, and has been the focus of studies on adaptation as a key element in genetic conservation of forests. Thumma et al. (2005) have used a candidate-gene-based LD mapping approach to identify alleles associated with microfibril angle, a wood quality trait affecting stiffness and strength of wood in eucalyptus. SNPs detected in the *Cinnamoyl Coa Reductase* gene, a key lignin gene, have been used to genotype 290 unrelated trees from a natural population of *E. nitens.* Several candidate genes affecting cell all biosynthesis in wood experiencing tension forces have been identified in eucalypts using microarray based global gene expression experiments (Qiu et al., 2008)

Thumma et al. (2009) have demonstrated the potential of revealing functional polymorphisms underlying quantitative traits by integrating both QTL and association mapping methods. A marker from the COBRA-like gene, whose Arabidopsis homologue has been implicated in cellulose deposition, was found to be strongly associated with a QTL for cellulose content in a full-sib family. By genotyping SNPs and a simple sequence repeat (SSR) marker in an association population, LD analysis has revealed that LD declines within the length of the COBRA- like gene. Subsequent association mapping analysis has contributed to fine-resolution mapping of the effect of this gene to a SNP marker.

As part of the Biotech MERCOSUR project 303 individuals from different open-pollinated progeny trials of *Eucalyptus globulus* core and intergrade populations were genotyped with the 7,680 DArT marker arrays and GWAS identified 16 markers that were associated with growth and two markers that were associated with lignin traits in (Cappa et al., 2013). Resende et al. (2017) used GWAS with the Regional Heritability Mapping (RHM) for the detection of true QTLs in 768 hybrid *Eucalyptus* trees, concluding that complex traits in *Eucalyptus* are controlled by multiple allele variants with rare effects. Very recently, Müller et al. (2019) carried out GWAS for growth traits with six single-marker models and regional heritability mapping (RHM) in four Eucalyptus breeding populations independently and by Joint-GWAS, using gene and segment-based models, with data for 3373 individuals genotyped with a communal EUChip60KSNP platform.

## VII. CONCLUSION

In Association Mapping, a large number of alleles present within the gene pool of a species are tested against the phenotype to detect significant associations. The genomic researchers are showing enormous interest in applying appropriate AM strategy to be followed for a given tree or plant is solely dependent on the patterns of LD and genomic resources available for a species. Due to high costs of sequencing and genotyping, marker development is considered very costly and thus only justifiable in crops of high commercial value. The important advantage of the rapid decay of LD is that once a marker–trait association has been discovered and validated, it is likely that such a marker is at a close physical distance to the functional variant probably within the gene itself or even is the functional variant itself. As the cost of genotyping is decreasing rapidly, researchers are shifting to association studies in forest trees for tree improvement programmes. There is an increasing trend of integrating gene expression data and even gene expression network information, attention is being paid to population size, and increasingly larger populations is being used to detect and map suitable markers.

The main question relies why only a few number of forest tree species is chosen for the association studies when there are hundreds of economically important species all over the world. Association mapping is well adapted in agricultural crops, but still not feasible or practical in majority of the forest tree species. Due to large undomesticated nature with prevalent outcrossing mating method, most of the forest tree population is highly heterozygous. In such populations racial effects or spatial relatedness between the trees affects the results of association mapping. As gene structure of most of the species in forest trees has not been mapped properly or reliably, the percentage of false positive associations is relatively high. This hinders the full potential of association studies to be used for tree improvement by the researchers. Currently, a number of researchers are working on association studies on forest trees as well as other plants in many laboratories worldwide. The near-future completion of genome sequencing projects of several species, powered with more cost-effective sequencing technologies, will certainly create a basis for application of whole genome-association studies, accounting for rare and common copy number variants (CNV) and epigenomics details of the trait of interest in plants. This will provide with more powerful association mapping tools for tree breeding and genomics programs in tagging true functional associations and consequently, marker based tree improvement. The examples of association studies performed in various plant germplasm resources largely depict the huge advancement of crop genomics era with the utilization of powerful LD-based association mapping tool which is also a good indicative of the potential utilization of this technology with other economically important tree species in the future.

## REFERENCES

[1] Abdurakhmonov, I.Y., & Abdukarimov, A. (2008). Application of Association Mapping to Understanding the Genetic Diversity of Plant Germplasm Resources. International Journal of Plant Genomics, vol. 2008, Article ID 574927, 18 pages. doi:10.1155/2008/574927
[2] Ball, R.D. (2005). Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. Genetics, 170(2): 859–873.
[3] Ball, R.D. (2007). Statistical analysis and experimental design. In: Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E., & de-Silva, H. N. (Eds.). Association Mapping in Plants (pp. 133– 196). New York: Springer.
[4] Beer, S.C., Siripoonwiwat, W., S.O'donoughue, L., Souza, E., Matthews, D., & Sorrels, M.E. (1997).

Associations between molecular markers and quantitative traits in an oat germplasm pool: can we infer linkages? Journal of Agricultural Genomics, vol. 3, paper 197.

[5] Cappa, E.P., El-Kassaby, Y.A., Garcia, M.N., Acun˜a, C., Borralho, N.M.G. et al. (2013). Impacts of Population Structure and Analytical Models in Genome-Wide Association Studies of Complex Traits in Forest Trees: A Case Study in Eucalyptus globulus. PLoS ONE, 8(11), Article ID: e81267. doi:10.1371/journal.pone.0081267

[6] Churchill, G.A., & Doerge, R.W. (1994). Empirical threshold values for quantitative trait mapping. Genetics, 138(3): 963–971.

[7] Descalsota, G., Swamy, B., Zaw, H., Inabangan-Asilo, M.A., Amparado, A., Mauleon, R., Reinke, R. (2018). Genome-Wide Association Mapping in a Rice MAGIC Plus Population Detects QTLs and Genes Useful for Bio fortification. Frontiers in plant science, 9: 1347. doi:10.3389/fpls.2018.01347

[8] Devlin, B., & Roeder, K. (1999). Genomic Control for Association Studies. Biometrics, 55: 997-1004. doi:10.1111/j.0006-341X.1999.00997.x.

[9] Dillon, S.K., Nolan, M., Li, W., Bell, C., Wu, H.X., Southertonet, S.G. (2010). Allelic variation in cell wall candidate genes affecting solid wood properties in natural populations and land races of Pinus radiate. Genetics, 185: 1477–1487.

[10] Eckert, A.J. et al. (2009b). High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (Pinus taeda, L.). Tree Genet. Genomes, 5: 225–234.

[11] Eckert, A.J., Bower, A.D., Wegrzyn, J.L., Pande, B., Jermstad, K.D., Krutovsky, K.V., Clair, J.B.S., Neale, D.B. (2009a). Association genetics of coastal Douglas-fir (Pseudotsuga menziesii var. menziesii, Pinaceae): I. Cold-hardiness related traits. Genetics, 182: 1289–1302.

[12] Ersoz, E.S., & Buckler, E.S. (2007). Applications of linkage disequilibrium and association mapping in crop plants. In: Varshney, R.K., & Tuberosa, R. (Eds.). Genomics-assisted crop improvement: vol. 1 (pp. 97–119). Genomics approaches and platforms, The Netherlands: Springer.

[13] Ersoz, E.S., Wright, M.H., Gonzalez-Martinez, S.C., Langley, C.H., Neale, D.B. (2010). Evolution of disease response genes in loblolly pine: insights from candidate genes, PLoS One, 5, Article ID: e14234.

[14] Falush, D., Stephens, M., & Pritchard, J. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics, 164: 1567-1587. doi:10.3410/f.1015548.197423

[15] Flint-Garcia, S., Thornsberry, J., Buckler, E. (2003). Structure of Linkage Disequilibrium in Plants. Annual review of plant biology, 54: 357-374. doi: 10.1146/annurev.arplant.54.031902.134907.

[16] Goldstein, D.B. & Weale, M.E. (2001). Population genomics: linkage disequilibrium holds the key. Current Biology, 11(14): R576–R579.

[17] Gonza´ lez–Martı´nez, S.C., Huber, D.A., Ersoz, E., Davis, J.M., Neale, D.B. (2008). Association genetics in Pinus taeda L. II. Carbon isotope discrimination. Heredity, 101: 19–26.

[18] Gonza´ lez–Martı´nez, S.C., Wheeler, N.C., Ersoz, E., Nelson, C.D., Neale, D.B. (2007). Association genetics in Pinus taeda L. I. Wood property traits. Genetics,175: 399–409.

[19] Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2014). Association Mapping in Crop Plants. Advances in Genetics, 109–147. doi:10.1016/b978-0-12-800271-1.00002-0

[20] Gupta, P. K., Rustgi, S., & Kulwal, P.L. (2005). Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Molecular Biology, 57(4): 461–485.

[21] Hall, D., Luquez, V., Garcia, M.V., St-Onge, K.R., Jansson, S., Ingvarsson, P.K. (2007). Adaptive population differentiation in phenology across a latitudinal gradient in European aspen (Populus tremula L.): a comparison of neutral markers, candidate genes and phenotypic traits. Evolution, 61: 2849–2860.

[22] Heuertz, M., De Paoli, E., Kallman, T., Larsson, H., Jurman, I., Morgante M., Lascoux, M., Gyllenstrand, N. (2006). Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [Picea abies (L.) Karst]. Genetics, 174: 2095–2105.

[23] Ingvarsson, P.K., & Street, N.R. (2011). Association genetics of complex traits in plants. New Phytologist, 189: 909-922. doi:10.1111/j.1469-8137.2010. 03593.x

[24] Ingvarsson, P.K., Garcia, M.V., Hall, D., Luquez, V., Jansson, S. (2006). Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (Populus tremula). Genetics, 172: 1845–1853.

[25] Ingvarsson, P.K., Garcia, M.V., Hall, D., Luquez, V., Jansson, S. (2008). Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (Populus tremula, Salicaceae). Genetics, 178: 2217–2226.

[26] Khan, M.A., & Korban, S.S. (2012). Association mapping in forest trees and fruit crops. Journal of Experimental Botany, 63(11): 4045–4060. doi:10.1093/jxb/ers105

[27] Kinloch, B.B.Jr., Parks, G.K., & Fowler, C.W. (1970). White pine blister rust: simply inherited resistance in sugar pine. Science, 167:193–195.

[28] Krutovsky, K.V., & Neale, D.B. (2005). Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. Genetics, 171: 2029–2041.

[29] Mackay, T. (2001). The genetic architecture of quantitative traits. Annual review of genetics, 35: 303-339. doi:10.1146/annurev.genet.35.102401.090633.

[30] Müller, B.S.F., de Almeida Filho, J.E., Lima, B.M., Garcia, C.C., Missiaggia, A., Aguiar, A.M., Takahashi, E., Kirst, M., Gezan, S.A., Silva-Junior, O.B., Neves, L.G., & Grattapaglia, D. (2019). Independent and Joint-GWAS for growth traits in Eucalyptus by assembling genome-wide data for 3373 individuals across four breeding populations. New Phytol, 221: 818-833. doi:10.1111/nph.15449

[31] Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E., Buckler, E.S. (2009). Association mapping: critical considerations shift from genotyping to experimental design. The Plant Cell, 21: 2194–2202.
[32] Neale, D.B. (2007). Genomics to tree breeding and forest health. Curr. Opin. Genet. Dev., 17: 539–544.
[33] Neale, D.B., & Kremer, A. (2011). Forest tree genomics: growing resources and applications. Nature Reviews Genetics, 12(2): 111–122. doi:10.1038/nrg2931
[34] Neale, D.B., & Savolainen, O. (2004). Association genetics of complex traits in conifers. Trends in Plant Science, 9: 325–330.
[35] Nordborg, M., Borevitz, J.O., Bergelson, J. et al. (2002). The extent of linkage disequilibrium in Arabidopsis thaliana. Nature Genetics, 30(2): 190–193.
[36] Oraguzie, N., Rikkerink, E., Gardiner, S., Silva, H. (2007). Association Mapping in Plants. doi:10.1007/978-0-387-36011-9
[37] Pavy, N., Paule, S., Parsons, L. et al. (2005). Generation, annotation, analysis and database integration of 16,500 white spruce EST clusters. BMC Genomics, 6(144).
[38] Pavy, N., Pegas, B., Beauseigle, S. et al. (2008). Enhancing genetic mapping of complex genomes through the design of highly- multiplexed SNP arrays: application to the large and un-sequenced genomes of white spruce and black spruce. BMC Genomics 9(21).
[39] Porth, I., Klapšte, J., Skyba, O., Hannemann, J., McKown, A. D., Guy, R.D. et al. (2013). Genome-wide association mapping for wood characteristics in Populus identifies an array of candidate single nucleotide polymorphisms. New Phytol, 200: 710–726. doi:10.1111/nph.12422
[40] Price, A., Patterson, N., Plenge, R., Weinblatt, M., Shadick, N., Reich, D. (2006). Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies. Nature genetics, 38: 904-909. doi:10.1038/ng1847
[41] Qiu, D., Wilson, I.W., Gan, S., Washusen, R., Moran, G.F., & Southerton, S.G. (2008). Gene expression in Eucalyptus branch wood with marked variation in cellulose microfibril orientation and lacking G-layers. New Phytol, 179(1): 94–103.
[42] Quesada, T., Gopal, V., Cumbie, W.P., Eckert, A.J., Wegrzyn, J.L., Neale, D.B., Goldfarb, B., Huber, D.A., Casella, G., Davis, J.M. (2010). Association mapping of quantitative disease resistance in a natural population of loblolly pine (Pinus taeda L.). Genetics, 186: 677–686.
[43] Rawat, A., Barthwal, S., Ginwal, H. (2014). Association mapping for resin yield in Pinus roxburghii Sarg. using microsatellite markers. Silvae Genetica, 63: 253-266. doi:10.1515/sg-2014-0033.
[44] Remington, D.L., Thornsberr, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M., Buckler, E.S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci, 98: 11479-11484.
[45] Resende, R.T., Resende, M.D., Silva, F.F., Azevedo, C.F., Takahashi, E.K., SilvaJunior, O.B. et al. (2017). Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in Eucalyptus. New Phytol, 213: 1287–1300. doi:10.1111/nph.14266
[46] Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. Science, 273: 1516–1517.
[47] Roy, J., Saugier, B., & Mooney, H. A. (Eds.). (2001). Terrestrial Global Productivity. Academic Press: London.
[48] Rungis, D., Hamberger, B., Berube, Y., Wilkin, J., Bohlmann, J., Ritland, K. (2005). Efficient genetic mapping of single nucleotide polymorphisms based upon DNA mismatch digestion. Molecular Breeding, 16: 261–270.
[49] Savolainen, O., & Pyhajarvi, T. (2007). Genomic diversity in forest trees. Current Opinion in Plant Biology, 10: 162–167.
[50] Singh, B.D., & Singh, A.K. (2015). Marker-Assisted Selection. In: Marker-Assisted Plant Breeding: Principles and Practices (pp. 217-256). New Delhi: Springer.
[51] Stich, B., Möhring, J., Piepho, H.P., Heckenberger, M., Buckler, E.S., & Melchinger, A.E. (2008). Comparison of mixed-model approaches for association mapping. Genetics, 178(3): 1745–1754. doi:10.1534/genetics.107.079707
[52] Thumma, B.R., Matheson, B.A., Zhang, D., Meeske, C., Meder, R., Downes, G.M., Southerton, S.G. (2009). Identification of a cis-acting regulatory polymorphism in a eucalyptus COBRA-like gene affecting cellulose content. Genetics, 183: 1153–1164.
[53] Thumma, B.R., Nolan, M.F., Evans, R., & Moran, G.F. (2005). Polymorphisms in Cinnamoyl CoA Reductase (CCR) are associated with variation in microfibril angle in Eucalyptus spp. Genetics, 171(3): 1257–1265.
[54] Tuskan, G.A. et al. (2006). The genome of black cottonwood Populus trichocarpa (Torr. & Gray). Science, 313: 1596–1604.
[55] UNEP. (2009). Vital Forest Graphics. United Nations Environmental Programme, 2009.
[56] Virk, P.S., Ford-Lloyd, B.V., Jackson, M.T., Pooni, H.S., Clemeno, T.P., & Newbury, H.J. (1996). Predicting quantitative variation within rice germplasm using molecular markers. Heredity, 76(3): 296–304.
[57] White, T.L., Adams, W.T., & Neale, D.B. (2007). Forest Genetics. CABI: Cambridge, MA.
[58] Yu, J., & Buckler, E.S. (2006). Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology, 17(2): 155-160. ISSN 0958-1669. doi:10.1016/j.copbio.2006.02.003
[59] Yu, J., Holland, J., McMullen, M., Buckler, E. (2008). Genetic Design and Statistical Power of Nested Association Mapping in Maize. Genetics, 178: 539-551. doi:10.1534/genetics.107.074245
[60] Yu, J., Pressoir, G., Briggs, W. et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet, 38: 203–208. doi:10.1038/ng1702

ISSN: 2394-4404

## APPENDIX: LIST OF ABBREVIATIONS USED

| | | | |
|---|---|---|---|
| AM | Association Mapping | MAGIC | Multiparent Advanced Generation Intercross |
| CCR | Cinnamoil CoA Reductase | MAS | Marker Assisted Selection |
| CNV | Copy Number Variants | MGA | Multi Gene Association |
| cpSSR | Chloroplast Simple Sequence Repeats | NAM | Nested Association Mapping |
| DH | Double Haploid | NGs | Next Generation Sequencing |
| ETS | Expressed sequence tags | NIL | Near Isogenic Line |
| GC | Genomic Control | nSSR | Nuclear Simple Sequence Repeats |
| GLD | Gametic Disequilibrium | QTL | Quantitative Trait Loci |
| GPD | Gametic Phase Disequilibrium | RFLP | Restriction Fragment Length Polymorphism |
| GSA | Genome Sequencing Annotation | RIL | Recombinant Inbred Line |
| GWAS | Genome Wide Association Studies | SA | Structured Association |
| HWE | Hardy Weinberg Equilibrium | SNP | Single Nucleotide Polymorphism |
| LD | Linkage Disequilibrium | SSR | Simple Sequence Repeats |
| MAB | Marker Assisted Breeding | UNEP | United Nations Environmental Programme |