Review Article

Molecular marker based genetic diversity in forest tree populations

Abstract

Information of the genetic diversity of the threatened tree species in any region of the world may contribute to the creation of effective strategies for their preservation and future use. Nowadays, molecular markers have proven to be invaluable tools for assessing genetic resources of tree plants by improving understanding of the users with regards to the distribution and the extent of genetic variation within and among species. Recently developed marker technologies allow the uncovering of the extent of the genetic variation in an unprecedented way through increased coverage of the genome. Markers have diverse applications in plant sciences, but certain marker types, due to their inherent characteristics, have also shown their limitations. A combination of diverse marker types is usually recommended to provide an accurate assessment of the extent of intra– and inter–population genetic diversity of naturally distributed plant species on which proper conservation directives for species that are at risk of decline can be issued. Here, specifically, natural populations of forest trees are reviewed by summarizing published reports in terms of the status of genetic variation in the pure species. In general, for out bred forest tree species, the genetic diversity within populations is larger than among populations of the same species, indicative of a negligible local spatial structure. Additionally, as is the case for plants in general, the diversity at the phenotypic level is also much larger than at the marker level, as selectively neutral markers are commonly used to capture the extent of genetic variation. However, more and more, nucleotide diversity within candidate genes underlying adaptive traits are studied for signatures of selection at single sites. This adaptive genetic diversity constitutes important potential for future forest management and conservation purposes.

Keywords: forest trees, genetic diversity, molecular markers

Temesgen Bedassa Gudeta
Department of Biology, Madda Walabu University, Ethiopia

Correspondence: Temesgen Bedassa Gudeta, Department of Biology, Madda Walabu University, P.O. Box 247, Robe, Ethiopia, Tel +2519-1178-5364, Email tasgabifenet@gmail.com

Introduction

Forest trees are largely undomesticated and highly heterozygous, due to their out crossing breeding systems and, therefore, have large effective population sizes.[1] Despite the high number of known species, approximately 450 different forest tree species are actively part of a deliberate domestication process through tree improvement programs (FAO).[2] Knowledge of the genetic diversity of the threatened tree species in any region of the world may contribute to the creation of effective strategies for their preservation and future use. The majority of the world–wide forests represent natural forests (93%), with 12% dedicated as conservation forests. A major concern regarding forests health and resilience is the declining in forest genetic diversity as documented as early as 1967 (FAO conference). Genetic diversity serves several important purposes: (a) as a resource for tree breeding and improvement programs to develop well–adapted tree species varieties and to enhance the genetic gain for a multitude of useful traits; (b) to ensure the vitality of forests as a whole by their capacity to withstand diverse biotic and abiotic stressors under changing and unpredictable environmental conditions; and (c) the livelihoods of indigenous and local communities that use traditional knowledge. Rich genetic diversity within and among forest tree species thus provides an important basis for maintaining food security and enabling sustainable development (FAO).[3]

Historically, for plant improvement, three major areas have always been important for molecular marker applications: (a) the determination of genetic diversity within, between and among populations; (b) verification and characterization of genotypes; and (c) marker–assisted selection (MAS).[4] In particular, for forest trees that are out crossing and largely undomesticated plant species, molecular markers have proven to be invaluable tools with applications in: (1) genetic conservation efforts by identification of genetic diversity hotspots; (2) the assembly of breeding populations in newly developed and advanced breeding programs; (3) the monitoring and characterization of population dynamics and gene flow; (4) the proper delineation of species taxonomy for management issues associated with conservation; (5) assessment of gene flow (pollen contamination) in seed orchards and the authentication of "controlled crossings", the assessment of inbreeding occurrence in breeding programs and studies of mating systems in non–industrial tree species; and (6) genetic fingerprinting in advanced breeding programs for the purpose of quality control to detect misidentified ramets in production and breeding populations.[4] Although tree breeding programs would significantly benefit from an early selection of clones with advantageous trait characteristics (particularly important for late–expressing wood quality traits), MAS was deemed not feasible for forest trees with limited genetic marker coverage.[5,6] The main reasons for the infeasibility of MAS as a tool for forest tree improvement are the inherent characteristics specific of forest trees as compared to inbred agricultural crop plants, such as the polygenic nature of most of the economically important traits in forestry, the inconsistency in quantitative trait locus (QTL) marker linkages among families originating from large outcrossed breeding populations and the instability of QTLs from the same genetic material planted across different sites, due to strong genotype–by–environment (G×E) interactions. As highly efficient next generation SNP (single nucleotide polymorphism) genotyping platforms have become available, genome–wide selection approaches have become feasible for accelerating forest tree breeding.[7,8]

Types of molecular markers and their applications

The use of DNA markers in plant and animal breeding has opened new territory in agriculture which is called molecular breeding. These markers are widely used because of their high prevalence and expression in different stages of the organisms.[9] This review is begun with regards to the genetic diversity in forest tree species with a brief historical retrospect concerning the development of marker types that have been widely employed for studying genetic variability in plants in general. The first, while the most easily accessible types of plant characteristics, are morphological markers that can easily be monitored based on simple inheritance.[9] However, due to serious drawbacks with respect to dominance, the difficulty of distinguishing between multiple alleles or even between different loci[10,11] and trait expression due to environmental and developmental variation (G × E interaction), their use was substantially reduced with the advent of DNA marker technologies. Another marker type that played an important role in assessing genetic diversity in plants was isozymes.[12,13] Isozymes had a long history in genetic variability studies in forestry, to assess the genetic diversity present within natural forest stands[14,15] or to determine whether domestication practices had led to a reduction in diversity.[16–18] However, the problem of these biochemical marker assays is that they are affected by plant phenological stage and their limited availability, and therefore, they would never allow for a genome–wide scan of variability (as only 0.1% of the total variation is detectable by this technique).[19] An invaluable alternative offered DNA–based markers, such as restriction fragment length polymorphism (RFLPs).[20–22] Finally, the possibility to rapidly amplify specific DNA fragments *in vitro* via polymerase chain reaction (PCR)[23] revolutionized the generation of molecular markers, leading to diverse sets of diagnostic DNA–marker systems with or without *a priori* sequence knowledge, such as random amplified polymorphic DNA (RAPD),[24] amplified fragment length polymorphism (AFLP),[25] simple sequence repeats (SSRs or microsatellites),[26] single nucleotide polymorphisms (SNPs)[27,28] and variations thereof.[29,30] For example, through bootstrap analysis the number of *loci* sufficient for the study of genetic diversity of *M. caesalpiniaefolia* (Figure 1) was estimated. In this review it is verified that, with the increase of loci analyzed in re–sampling there was an increase of the values of correlation and a reduction of the Kruskal stress values.
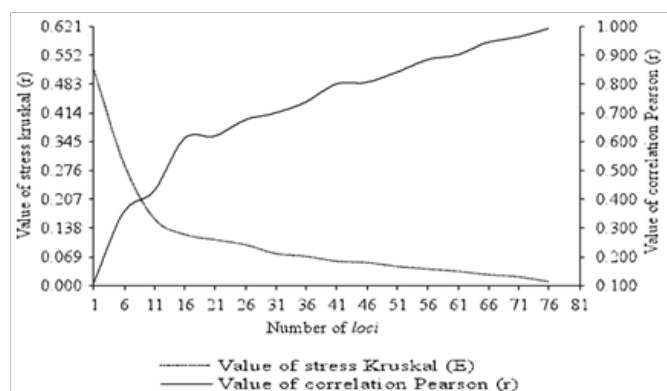

Figure 1 Values of the Pearson correlation (r) and Kruskal stress (E) as a function of the number of ISSR loci used to estimate the genetic diversity of nine *M. caesalpiniaefolia* individuals.[36]

Important issues are related to the reproducibility of the RAPD marker system,[31] other limitations, such as the presence of null alleles

in the case of SSR assays that may underestimate heterozygosity,[32] or the dominance nature of the RAPD and AFLP marker systems, where heterozygous individuals cannot be distinguished from homozygous ones, and lastly, the inexpensive generation of a vast abundance of highly polymorphic DNA markers to tackle genome–wide genetic diversity studies. Dependent on the study focus, genetic markers were derived from nuclear or organelle sequences; for example, chloroplast–or mitochondrial–derived diagnostic markers,[33–35] dependent on the evidence of their maternal inheritance in the species, were used to trace back the colonization history of angiosperm forest tree species and conifers, respectively.[36,37] Although it has been known that variability within protein–coding regions is far less than within non–coding genomic regions, due to lower mutation rates and purifying selection to maintain proper protein functions, the study of polymorphic sites within coding sequences has been deemed more relevant because of their putative functional associations and, in addition, the ease of their interspecific transferability for comparative genetic studies based on sequence conservation. Thus, a major focus in plant studies has been the development of genetic markers prevalently present within such coding regions for high–throughput analysis of many samples using the inexpensive detection method of PCR fragment length polymorphisms (e.g., eco–tilling to circumvent expensive Sanger resequencing of PCR products, as in the case of SNP detection and genotyping),[38] but that still relied on laborious PCR optimizations.[39–42] The substantial and almost exponential drop in whole genome sequencing costs, thanks to the 1000 Human Genome Project, which has stimulated the development of highly cost–efficient high–throughput technologies, has also provided for the plant research community unprecedented opportunities for affordable in–depth characterization of plant genomes that has involved the genome–wide discovery of SSRs and SNPs and the detection of common, as well as rare functional variants by next generation sequencing.[43–49]

Advantages and applications of molecular markers in plants

i) Time saving: genomic DNA can be isolated from any part of the plant tissue at every stage of its development and target trait information can be obtained with linked DNA markers before pollination, thus allowing breeders to carry out more informed genetic crosses.[9,50]

ii) Stability and reliability: phenotypic evaluation of genetic traits is often complicated by environmental factors. However, DNA markers are mostly neutral to environmental variation. The breeder can evaluate their material independently of the environmental conditions, environmental conditions can be favourable or unfavourable for morphologic and/or biochemical marker expression.[9]

iii) Biosafety: diagnostic tests for the presence or absence of traits for disease resistance can be conducted by DNA markers tightly linked to the target gene without resorting to pathogen inoculation in the field or greenhouse. Molecular markers also facilitate introgression of genes into elite cultivars in advance of the occurrence of certain races of diseases or biotypes of insects.

iv) Performance: evaluation of breeding lines in early generations of the breeding process with DNA markers can allow breeders to reject progenies from the programme and improve the genetic quality of breeding materials.

v) Precise selection of the complex traits: polygenic traits are often

difficult to select for using conventional breeding approaches. DNA markers linked to quantitative trait loci (QTL) allow them to be treated as single Mendelian factors. Beside analyzing and selecting the interesting characters, molecular markers allow the researchers also to analyze the wild species with potential interest for the breeding program.[51] The structure of genetic linkage maps using molecular markers is based on certain principles: a) selecting the molecular markers and genotyping system, b) selecting the parental lines from germplasm collections which are highly polymorphic in the marker loci, c) creating populations or lines (derived from these populations) using a large number of molecular markers segregating in the population (Figure 2), d) genotyping of each individual/ line using molecular markers and making linkage maps using markers information.

vi) In the case of Marker assisted pyramiding: Pyramiding is the simultaneous integration of multiple genes/ QTLs into a single genotype. The most widely application of pyramiding is the integration of multiple disease resistance genes into a plant for durable resistance to a disease. The main advantage of molecular markers in gene pyramiding is their ability to search and discover multiple genes in plants whose phenotypic effects are difficult to be separated.[9,51]
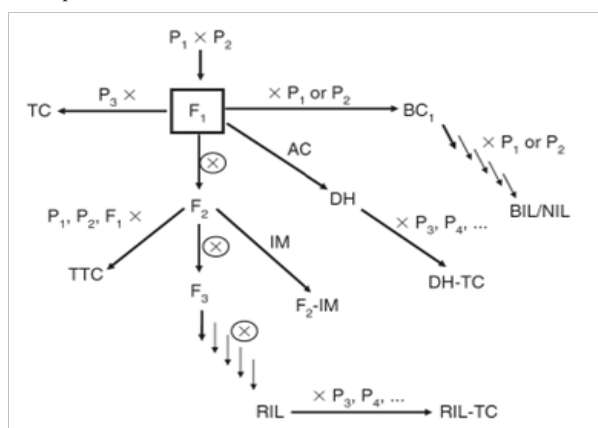

Figure 2 Examples of mapping populations and their relationship.[51] AC, anther culture; BC, backcross population; BIL, backcross inbred line; DH, double haploid; IM, intermating; NIL, near–isogenic line; RIL, recombinant inbred line; TC, testcross; TTC, triple testcross.
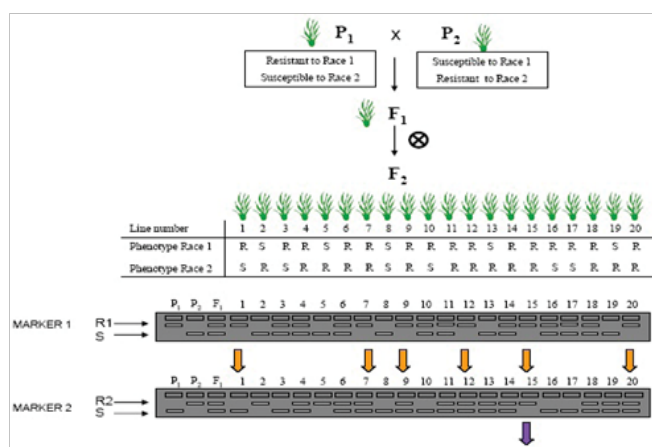

Figure 3 Marker assisted pyramiding of two disease resistance genes. Note that homozygotes can be selected from the F_2 population.[9]

The most widely application of pyramiding is the integration of several genes for disease resistance (i.e., integration of qualitative resistance genes) into a single genotype. The motivation of this work is to develop "durable" or stable resistance to a disease, because pathogens usually overcome single–gene resistance over time due to the emergence of new strains of plant pathogens.

Evaluation of genetic diversity

A number of evolutionary processes can impact the genetic diversity of natural populations. These are: (a) spontaneously arising mutations; (b) gene flow via migration; (c) inbreeding; (d) natural selection; (e) the Wahlund effect; and (f) random genetic drift.[50] Genetic drift introduces random changes in allele frequencies over generations and becomes important for finite population samples and/or a large number of generations. These random allele frequency changes can, over time, lead to allele fixation or extinction. By all means, genetic drift represents a source of differences in genetic diversity among different populations. On the other hand, gene flow evens out among–population genetic differences, but increases genetic variation within populations, due to the introduction of new alleles. Selection influences within–population diversity, but the effects are dependent on the nature of these selection processes (balancing selection). Furthermore, the effects of natural selection are interwoven with stochastic effects, such as genetic drift. Mutations can counterbalance the loss of allelic diversity; however, natural mutations are rare, and such mutations that turn out to be harmful allelic variants are again removed by purifying selection. The occurrence of a population bottleneck causes a significant reduction in the effective population size and represents a major reason for the loss in allelic diversity, first by the loss of rare alleles, then by the successive loss of heterozygosity in the population.[50] Inbreeding and the presence of a subpopulation structure, where gene flow is prevented by habitat fragmentation (the Wahlund effect), both cause the loss in heterozygosity.[50] This, in turn, results in increased genetic diversity among populations.

Within–population genetic variation using genotype data

A gene is defined as polymorphic in the population when its most common allele is less frequent than 95%.[50] Genetic diversity can be assessed by estimating the following parameters: the total number of different alleles in the population, the percentage of polymorphic loci, the mean number of alleles per locus, the allelic richness, the within–population genetic diversity, θ, the effective population size, Ne (*i.e.*, θ divided by the per–generation mutation rate), the minor allele frequency (as in the case of biallelic loci), the proportion of heterozygous individuals in the population for a given locus (the expected heterozygosity, (HE; based on the Hardy–Weinberg expectations that assume the random mating of genotypes), as well as the observed heterozygosity (H_O) and the fixation index, F.[50] Genomic diversity is estimated by genome–wide assessment of genetic diversity using a larger sample of loci at random. An estimate of the genome–wide genetic diversity in a population is then derived by averaging heterozygosity over the multitude of studied loci.

Between–/among–population genetic variation using genotype data

Differences in the genetic diversity between/among (sub–) populations are assessed based on the presence of significant allele frequency differences; widely applied metrics to estimate such

"genetic differentiation" include, for example, F_{ST}[51,52], θ [53], R_{ST}[54], Φ_{ST} (Φ'_{ST})[55,56], G_{ST}(G'_{ST})[57,58], D_{ST}[57], H_{ST}[59] or D.[60] Some measures are marker–dependent; they are based on the assumption of infinite–allele or stepwise mutation models, respectively, and depending on whether biallelic or multi–allelic molecular markers or haplotype data were used in the analysis (F_{ST}; R_{ST}; Φ ST). Moreover, the use of fixation measures for result interpretation with regard to genetic differentiation has been found to be problematic when the populations under study exhibited high genetic diversity/heterozygosity.[58,60] For such cases, "standardized" genetic differentiation metrics' have been suggested;[56,58,60] but, see also the recent publication on the topic by Whitlock et al.[61] who emphasized the continuous use of F_{ST} for intra–specific differentiation estimation when the mutation rate is small (relative to gene flow), while emphasizing the use of Φ_{ST} and R_{ST} when the mutation rate is high (as in the case of SSRs). In any case, for the estimation of population divergence from genotypic data, freely available software packages within the R environment[62] that have these statistics implemented are readily available (*cf.* "mmod"). Furthermore, genetic loci with allelic frequencies significantly different among populations and potentially under selection ("F_{ST} outlier loci") can be efficiently detected using multilocus scans that compare the patterns of nucleotide diversity and genetic differentiation (based on the distribution of empirical F_{ST} estimates conditioned on H_E) to the simulated genome–wide selectively–neutral genetic background.[63,64]

Sequence divergence using sequence alignment data

Other and additional ways to look at genetic diversity and study mutation and selection events within populations and by comparing different populations involve the characterization of DNA sequences of genes and the diversity of nucleotides as the specific study entities.[65–68] Widely used tests include nucleotide diversity π ,[50,69–72] and the McDonald–Kreitman & HKA (Hudson–Kreitman–Aguade) tests,[73,74] respectively. Such tests are implemented in the freely available software package, DnaSP.[75] The combination of results from such analyses has particular value for identifying past population size changes (population expansion or population bottleneck).

Diversity of forest tree population

One of the first comprehensive reviews on genetic diversity with regards to forest tree populations was published by Hamrick et al.[76] This early work summarized results based on isozymes and is especially valuable, as it compares long–lived forest trees with other life forms of plant species, in total comprising 662 different species with representatively high sample sizes for the analysis of the genetic diversity parameters. Long–lived, woody species showed the highest genetic diversity (including a significantly higher percentage of polymorphic loci and more alleles per locus) among all plant species. Specifically, the genetic diversity within populations was significantly the highest (HE=0.15) compared to all other plant life forms (HE<0.10). However, heterogeneity in genetic diversity exists among woody species taxa and this is due to the different evolutionary histories of species. For example, species from smaller founder populations, small disjunct populations or those with past population bottlenecks show generally less genetic diversity. *Alseis blackiana*, *Picea glauca*, *Robinia pseudoacacia* and *Pinus sylvestris* showed high diversity. On the other side of the spectrum were *Acacia mangium*, *Pinus resinosa*, *P. torreyana* and *Populus balsamea* with very low diversity.[76] Other studies[77,78] identified additional species with low intra–population diversity: *Ficus carica* and *Thuja plicata*.

While most studies identified high intra–population variation, by contrast, the diversity among populations of long–lived, woody tree species based on the GST estimate was significantly the lowest (GST =0.08) compared to the herbaceous and annual life forms (GST>0.25).[76] When woody angiosperms were compared to gymnosperms in terms of their intra–population genetic diversity, differences were not significant, yet the latter exhibited a significantly higher percentage of polymorphic allozyme loci, suggestive of a higher proportion of low frequency alleles in gymnosperm species.[76] Angiosperm species showed higher among–population genetic diversity (GST). Recent research on the conifer genome evolution, which involved orthologous coding sequence alignments for thousands of gymnosperms and angiosperm orthologous coding sequences, respectively, showed, more specifically, an overrepresentation of non–synonymous substitutions in protein–coding genes for conifers compared to angiosperms,[79] while the average synonymous mutation rate in angiosperms is significantly higher, suggestive of a higher number of fixed adaptive mutations in conifers. As expected, the extent of the geographical range had a significant impact on genetic diversity within species and among populations.[76] Geographically widespread species showed a significantly higher intra–population genetic diversity estimate compared to locally confined species, but the latter showed higher genetic diversity among populations.[76] However, the "non–significant" inter–population differentiation sometimes reported in these isozyme studies (see above) can mislead the directions of conservation efforts. Other marker types, those that are able to cover a higher portion of the overall genetic variation (such as restriction fragment length polymorphisms of DNA) succeeded in uncovering significant among–population diversity in *Pinus* and *Quercus*, specifically with the application of organellar DNA markers.[80,81] Differing outcomes for isozymes and organellar DNA studies on population divergence are frequent and were even reported within the same sample as for *Argania spinosa* (L.) Skeels, an important multi–purpose tree in the Moroccan local community.[82] It is also clear that variation at selectively neutral molecular markers commonly used to assess genetic diversity within or among populations may not covary with the phenotypic expression of a particular qualitative or quantitative trait of interest,[29] such that population differentiation for adaptive traits (growth, morphology or fitness) is much higher than for isozymes, for example. In any case, the total allelic richness was identified as a more adequate directive than the HE estimate for conservation purposes, and marker types, such as SSRs or DNA sequence–based data, that are highly polymorphic are required for an accurate estimate.[82] A recent study integrating molecular genetic analysis based on four SSR and five sequence loci along with climate modeling[83] forecasted the long–term decline of the late–successional Australian rainforest conifer, *Podocarpus elatus*, in its southern populations, due to habitat fragmentation (and the decline in Ne), for which conservation strategies are now invoked. Isozyme markers (15 loci) were used to characterize the genetic diversity of *Carapa procera*, which occurs in low density within a tropical rain forest.[15] Its characteristics were high within–population diversity (comparable to temperate gymnosperms), high heterozygosity and a lack of spatial structure consistent with the highly outcrossing nature of the species, leading to extensive pollen–mediated gene flow that prevented local genetic differentiation. When 63 SNP polymorphisms (surveyed by eco–tilling) in nine different genes with broad functional properties were targeted as a feature for understanding DNA variation in 41 wild populations of a small western black cottonwood (*P. trichocarpa*) sample panel,[40] it was found that heterozygosity was high (HO =0.47) and that overall nucleotide diversity at the gene level (π=0.0018)

among populations was low. Similarly, low average π values of the segregating sites were obtained for other forest tree species, such as *P. nigra* (π=0.0024)[84] and *Pinus sylvestris* (π=0.0025).[85] Much higher overall nucleotide diversity levels in a conifer were uncovered for *P. taeda* (π=0.00398).[86] Among the studied poplars, interestingly, the European species, *P. tremula*, showed the highest nucleotide diversity (π=0.007 or even π=0.0111,[87,88] dependent on the surveyed genes), but differences in diversity were also consistent with its different and complex demographic history. However, nucleotide diversity is best interpreted on a gene–by–gene basis, as population history and selection affect these mutation rates more specifically.[40,89] In a similar context, assessing the adaptive genetic diversity in forest trees is important to harness this adaptive potential for future forest management and conservation purposes.[90] Candidate genes underlying a specific trait of interest are typically selected (*cf.* nine candidate genes for bud burst in *Quercus petraea*: π=0.00615;[91] 121 candidate genes for cold hardiness in *Pseudotsuga menziesii* var. *menziesii*: π =0.004;[92] 13 candidate genes for drought stress in *Pinus pinaster* π=0.00548).[64] While most of this detected variation was largely attributed to purifying selection (an excess of nucleotide diversity at synonymous *vs.* non–synonymous sites), as commonly observed in forest trees, patterns of strong diversifying selection in candidate genes were also uncovered.[64]

Conclusion

This review is focused on the value of integrating knowledge on adaptive complex traits as a companion to molecular markers for making informative management and conservation decisions. It is emphasized that integrative approaches using future climate modeling have been very successful in uncovering potential threats of declines of the genetic diversity and the distribution of forest tree species, so that timely precautions to preserve the species can be undertaken. Associated with the substantial drop in whole genome sequencing costs making the sequencing of genetically complex organisms more affordable, inventorying the complete portfolio of genetic resources has become feasible. This will also open new avenues for the conservation of previously marginalized and undervalued forest tree species that are considered of less economic value, but nevertheless represent value to the local ecosystems. While the present review focused primarily on the genetic diversity assessed for pure species, it is also stressed on the importance of investigating natural species hybrid zones as important sources of population genetic diversity in forest tree management. The potential of molecular markers application for the management forest gene conservation is summarized as following:

a) To clarify the identity of the taxa and their relatedness as well as infer their evolutionary histories

b) To rectify the correct clones and ramets in gene banks to avoid mislabeling, duplication and contamination.

c) To evaluate the amount, extent and distribution of genetic variation within and among populations

d) To estimate mating system (selfing and out–crossing rate) and gene flow.

e) To evaluate the status of genetic resources as the criteria for *ex situ* and *in situ* conservation from genetic information provided.

f) To maximize the management of gene conservation by combining adaptive traits, ecogeograhic and genetic survey for both *ex situ* collecting programs and for identifying sites for *in situ* conservation.

Therefore, how to manage, which and how many materials we should manipulate and where we should establish or protect the gene resources, depends on whether we really know the genetic background of particular species which we wants to conserve.

Acknowledgements

None.

Conflict of interest

Author declares there is no conflict of interest.

References

1. Jansson S, Ingvarsson PK. Cohort–structured tree populations. *Heredity.* 2010;105:331–332.

2. Food and Agriulture Organization of the United Nations. *Report of the 14th Regular Session of the Commission on Genetic Resources for Food and Agriculture.* Italy: FAO; 2017. p. 1–81.

3. Haines R. *Biotechnology in Forest Tree Improvement: With Special Reference to Developing Countries.* 1st ed. Italy: Food and Agriculture Organization of the United Nations; 1994. p. 230.

4. Lapinjoki S, Tammisola J, Kauppinen V, et al. *The Finnish project for genetic mapping and its applications in breeding, biotechnology and research of white birch.* USA: In Proceedings of the Sixth Meeting of the International Conifer Biotechnology Working Group; 1992.

5. Strauss SH, Lande R, Namkoong G. Limitations of molecular–marker–aided selection in forest tree breeding. *Canadian Journal of Forest Research.* 1992;22(7):1050–1061.

6. Grattapaglia D, Resende MDV. Genomic selection in forest tree breeding. Tree Genet. *Genomes.* 2011;7(2):241–255.

7. Resende MDV, Resende MFR, Sansaloni CP, et al. Genomic selection for growth and wood quality in Eucalyptus: Capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist.* 2012;194(1):116–128.

8. Kordrostami M, Rahimi Mihdi. Molecular markers in Plants: Concepts and applications. *G3M.* 2015;3(5):4024–4031.

9. Jain, SK, Allard RW. Population studies in predominantly self–pollinated species, 1. Evidence for heterozygote advantage in a closed population of barley. *USA: PNAS;* 1960;46(10):1371–1377.

10. Brown AHD. Isozymes, plant population genetic structure and genetic conservation. *Theoretical and Applied Genetics.* 1978;52(4):145–157.

11. Andersen JR, Lubberstedt T. Functional markers in plants. *Trends in Plant Science.* 2003;8(11):554–560.

12. Brown AHD. Enzyme polymorphism in plant–populations. *Theoretical Population Biology.* 1979;15(1):1–42.

13. Ritland K, Jain SA. A comparative–study of floral and electrophoretic variation with life–history variation in *Limnanthes–alba* (Limnanthaceae). *Oecologia.* 1984;63(2):243–251.

14. Ritland K, Meagher LD, Edwards DGW, et al. Isozyme variation and the conservation genetics of Garry oak. *Canadian Journal of Botany.* 2005;83(11):1478–1487.

15. Doligez A, Joly HI. Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (*Meliaceae*), in French Guiana. *Heredity.* 1997;79:72–82.

16. El–Kassaby Y, Ritland K. Impact of selection and breeding on the genetic diversity in Douglas–fir. *Biodiversity & Conservation.* 1996;5(6):795–813.

17. El–Kassaby YA. Domestication and genetic diversity—Should we be concerned? *The Forestry Chronicle.* 1992;68(5):687–700.

18. Chaisurisri K, El–Kassaby YA. Genetic diversity in a seed production population vs. natural–populations of sitka spruce. *Biodiversity & Conservation.* 1994;3(6):512–523.

19. El–Kassaby YA. Genetic variation within and among conifer populations: Review and evaluation of methods. In *Biochemical Markers in the Population Genetics of Forest Trees.* In: Fineschi S, Malvolti ME, Cannata F, Hattemer HH, editors. The Netherlands: SPB Academic Publishing; 1991. p. 61–76.

20. Botstein D, White RL, Skolnick M, et al. Construction of a genetic–linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics.* 1980;32(3):314–3311.

21. Szmidt AE, El–Kassaby YA, Sigurgeirsson A, et al. Classifying seedlots of *Picea–sitchensis* and *Picea–glauca* in zones of introgression using restriction analysis of chloroplast DNA. *Theoretical and Applied Genetics.* 1988;76(6):841–845.

22. Glaubitz JC, El–Kassaby YA, Carlson JE. Nuclear restriction fragment length polymorphism analysis of genetic diversity in western redcedar. *Canadian Journal of Forest Research.* 2000;30(3):379–389.

23. Mullis K, Faloona F, Scharf S, et al. Specific enzymatic amplification of DNA *in vitro*—The polymerase chain–reaction. *Cold Spring Harbor Symposia on Quantitative Biology.* 1986;51:263–273.

24. Williams JGK, Kubelik AR, Livak KJ, et al. DNA polymorphisms amplified by arbitrary primers are useful as genetic–markers. *Nucleic Acids Research.* 1990;18(22):6531–6535.

25. Vos P, Hogers R, Bleeker M, et al. AFLP—A new technique for DNA–fingerprinting. *Nucleic Acids Research.* 1995;23(21):4407–4414.

26. Zietkiewicz E, Rafalski A, Labuda D. Genome fingerprinting by simple sequence repeat (SSR)–anchored polymerase chain–reaction amplification. *Genomics.* 1994;20(2):176–183.

27. Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology.* 2002;5(2)94–100.

28. Buckler ES, Thornsberry JM. Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology.* 2002;5(2):107–111.

29. Mondini L, Noorani A, Pagnotta M. Assessing plant genetic diversity by molecular tools. *Diversity.* 2009;1(1):19–35.

30. Nybom H, Weising K, Rotter B. DNA fingerprinting in botany: Past, present, future. *Investigative Genetics.* 2014;5(1):1–35.

31. Perez T, Albornoz J, Dominguez A. An evaluation of RAPD fragments reproducibility and nature. *Molecular Ecology.* 1998;7(10):1347–1357.

32. Callen DF, Thompson AD, Shen Y, et al. Incidence and origin of "null" alleles in the (AC)n microsatellite markers. American *Journal of Human Genetics.* 1993;52(5):922–927.

33. Sutton BCS, Flanagan DJ, El–Kassaby YA. A simple and rapid method for estimating representation of species in spruce seedlots using chloroplast DNA restriction–fragment–length– polymorphism. *Silvae Genetica.* 1991;40:119–123.

34. Sutton BCS, Flanagan DJ, Gawley JR, et al. Inheritance of chloroplast and mitochondrial–dna in Picea and composition of hybrids from introgression zones. *Theoretical and Applied Genetics.* 1991;82(2):242–248.

35. Demesure B, Sodzi N, Petit RJ. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast dna in plants. *Molecular Ecology.* 1995;4(1):129–131.

36. Fernando dos SA, Vasconcelos MP, Fábio V, et al. ISSR molecular markers for the study of the genetic diversity of *Mimosa caesalpiniaefolia* Benth. *IDESIA.* 2016;34(3):47–52.

37. Petit RJ, Csaikl UM, Bordacs S, et al. Chloroplast DNA variation in European white oaks—Phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management.* 2002;156(1–3):5–26.

38. Gamache I, Jaramillo–Correa JP, Payette S, et al. Diverging patterns of mitochondrial and nuclear DNA diversity in subarctic black spruce: Imprint of a founder effect associated with postglacial colonization. *Molecular Ecology.* 2003;12(4):891–901.

39. Comai L, Young K, Till BJ, et al. Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *The Plant Journal.* 2004;37(5):778–786.

40. Parida SK, Kumar KAR, Dalal V, et al. Unigene derived microsatellite markers for the cereal genomes. *Theoretical and Applied Genetics.* 2006;112(5):808–817.

41. Gilchrist EJ, Haughn GW, Ying CC, et al. Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of *Populus trichocarpa. Molecular Ecology.* 2006;15(5):1367–1378.

42. Liewlaksaneeyanawin C, Zhuang J, Tang M, et al. Identification of COS markers in the Pinaceae. *Tree Genetics & Genomes.* 2009;5(1);247–255.

43. Bodénès C, Chancerel E, Gailing O, et al. Comparative mapping in the Fagaceae and beyond with EST–SSRs. *BMC Plant Biology.* 2012;12:153–170.

44. Poland JA, Rife TW. Genotyping–by–sequencing for plant breeding and genetics. *Plant Genome.* 2012;5(3):92–102.

45. Davey JW, Hohenlohe PA, Etter PD, et al. Genome–wide genetic marker discovery and genotyping using next–generation sequencing. *Nature Reviews Genetics.* 2011;12:499–510.

46. Zalapa JE, Cuevas H, Zhu HY, et al. Using next–generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany.* 2012;99(2):193–208.

47. Geraldes A, Pang J, Thiessen N, et al. SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing. *Molecular Ecology Resources.* 2011;11(S1):81–92.

48. Chen C, Mitchell S, Elshire R, et al. Mining conifers' mega–genome using rapid and efficient multiplexed high–throughput genotyping–by–sequencing (GBS) SNP discovery platform. *Tree Genetics & Genomes.* 2013;9(6):1537–1544.

49. Tsai H, Howell T, Nitcher R, et al. Discovery of rare mutations in populations: TILLING by sequencing. *Plant Physiology.* 2011;156:1257–1268.

50. Falque M, Santoni S. *Molecular markers and high–throughput genotyping analysis.* In: Morot–Gaudry JF, Lea P, Briat JF, editors. *Functional plant genomics.* UK: Taylor & Francis Group; 2007. p. 503–527.

51. MacKill DJ. *Breeding for resistance to abiotic stresses in rice: the value of quantitative trait loci. Plant breeding: The Arnel R Hallauer International Symposium.* USA: Wiley Online Library; 2006.

52. Marroni F, Pinosio S, Di Centa E, et al. Large–scale detection of rare variants via pooled multiplexed next–generation sequencing: Towards next–generation Ecotilling. *The Plant Journal.* 2011;67(4):736–745.

53. Hartl DL, Clark AG. *Principles of Population Genetics.* 4th ed. USA: Sinauer Associates Inc; 2007. p. 652.

54. Wright S. *Evolution and the Genetics of Populations, Volume 2: The Theory of Gene Frequencies.* USA: University of Chicago Press; 1984.

55. Raymond M, Rousset F. An exact test for population differentiation. *Evolution*. 1995;49(6):1280–1283.

56. Weir BS, Cockerham CC. Estimating f–statistics for the analysis of population–structure. *Evolution*. 1984;38(6):1358–1370.

57. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*. 1995;139(1):1463–1463.

58. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes—Application to human mitochondrial–DNA restriction data. *Genetics*. 1992;131(2):479–491.

59. Meirmans PG. Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*. 2006;60(11):2399–2402.

60. Nei M. Analysis of gene diversity in subdivided populations. *USA: PANS*. 1973;70(12):3321–3323.

61. Hedrick PW. A standardized genetic differentiation measure. *Evolution*. 2005;59(8):1633–1638.

62. Lewontin RC. The apportionment of human diversity. In: Dobzhansky T, Hecht MK, Steere WC, editors. USA: Appleton–Century–Crofts; 1972. p. 381–398.

63. Jost L. G(ST) and its relatives do not measure differentiation. *Molecular Ecology*. 2008;17(18):4015–4026.

64. Whitlock MC. G'(ST) and D do not replace F(ST). *Molecular Ecology*. 2011;20(6):1083–1091.

65. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Austria: R Foundation for Statistical Computing; 2008.

66. Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*. 1996;263(1377):1619–1626.

67. Eveno E, Collada C, Guevara MA, et al. Contrasting patterns of selection at Pinus pinaster Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution*. 2008;25(2):417–437.

68. Kimura M. Number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*. 1969;61(4):893–903.

69. Ohta T. Mutational pressure as main cause of molecular evolution and polymorphism. *Nature*. 1974;252(5482):351–354.

70. Tajima F. Evolutionary relationship of DNA–sequences in finite populations. *Genetics*. 1983;105(2):437–460.

71. Kreitman ME, Aguade M. Excess polymorphism at the *Adh locus* in *Drosophila–melanogaster*. *Genetics*. 1986;114(1):93–110.

72. Tajima F. Statistical–method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123(3):585–595.

73. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993;133(3):693–709.

74. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics*. 2000;155(3):1405–1413.

75. Zeng K, Fu YX, Shi S, et al. Statistical tests for detecting positive selection by utilizing high–frequency variants. *Genetics*. 2006;174(3):1431–1439.

76. McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh locus* in Drosophila. *Nature*. 1991;351(6328):652–654.

77. Hudson RR, Kreitman M, Aguade M. A test of neutral molecular evolution based on nucleotide data. *Genetics*. 1987;116(1):153–159.

78. Rozas J, Sanchez–DelBarrio JC, Messeguer X, et al. DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. 2003;19(18):2496–2497.

79. Hamrick JL, Godt MJW, Sherman–Broyles SL. Factors influencing levels of genetic diversity in woody plant species. *Population Genetics of Forest Trees*. 1992;6:95–124.

80. Mueller–Starck G, Baradat P, Bergmann F. Genetic variation within European tree species. *New Forest*. 1992;6(1–4):23–47.

81. Neale DB, Williams CG. Restriction–fragment–length–polymorphism mapping in conifers and applications to forest genetics and tree improvement. *Canadian Journal of Forest Research*. 1991;21(5):545–554.

82. Buschiazzo E, Ritland C, Bohlmann J, et al. Slow but not low: Genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evolutionary Biology*. 2012;12:8–21.

83. Strauss SH, Hong YP, Hipkins VD. High–levels of population differentiation for mitochondrial–DNA haplotypes in *Pinus–radiata*, *muricata*, and *attenuata*. *Theoretical and Applied Genetics*. 1993;86:605–611.

84. Kremer A, Petit RJ, Zanetto A, et al. Nuclear and organelle gene diversity in *Quercus robur* and *Q. petraea*. *Genetic Variation in European Populations of Forest Trees*. In: Müller–Starck G, Ziehe M, editors. Germany: Sauerländer's Verlag; 1991. p. 141–166.

85. Petit RJ, El Mousadik A, Pons O. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*. 1998;12(4):844–855.

86. Mellick R, Rossetto M, Allen C, et al. Intraspecific divergence associated with a biogeographic barrier and climatic models show future threats and long–term decline of a rainforest conifer. *The Open Conservation Biology Journal*. 2013;7:1–10.

87. Zaina G, Morgante M. *Nucleotide Diversity and Linkage Disequilibrium in Populus nigra*. Italy: In Proceedings of the XLVIII Italian Society of Agricultural Genetics—SIFV–SIGA Joint Meeting; 2004.

88. Garcia–Gil MR, Mikkonen M, Savolainen O. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Molecular Ecology*. 2003;12(5):1195–1206.

89. Brown GR, Gill GP, Kuntz RJ, et al. Nucleotide diversity and linkage disequilibrium in loblolly pine. *USA: PNAS*. 2004;101(42):15255–15260.

90. Ismail M, Soolanayakanahally RY, Ingvarsson PK, et al. Comparative nucleotide diversity across North American and European populus species. *Journal of Molecular Evolution*. 2012;74(5–6):257–272.

91. Ingvarsson PK. Nucleotide polymorphism and linkage disequilbrium within and among natural populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics*. 2005;169(2):945–953.

92. Mosca E, Eckert AJ, Liechty JD, et al. Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evolutionary Applications*. 2012;5(7):762–775.