Putative local adaptive SNPs in the genus Avicennia

Laleh Malekmohammadi
Shahid Beheshti University
Masoud Sheidai (msheidai@sbu.ac.ir)
Shahid Beheshti University
Farrokh Ghahremaninejad
Kharazmi University
Afshin Danehkar
University of Tehran
Fahimeh Koohdar
Shahid Beheshti University

Research Article

Keywords: Avicennia, Geographical adaptation, CCA, Manhattan plot, RDA, SNPs

Posted Date: March 29th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1488831/v1

Version of Record: A version of this preprint was published at Biochemical Genetics on April 3rd, 2023.
See the published version at https://doi.org/10.1007/s10528-023-10362-4.

Abstract

The genus *Avicennia* with eight species grow in intertidal zones of tropical and temperate regions, ranging in distribution from West Asia, to Australia, and Latin America. These mangroves have several medicinal applications for mankind. Many genetic and phylogenetic studies have been carried out on mangroves, but none is concerned with geographical adaptation of SNPs. We therefore, used ITS sequences of about 120 *Avicennia* taxa growing in different parts of the world and undertook computational analyses to identify discriminating SNPs among these species and to study their association with geographical variables. A combination of multivariate and Bayesian approaches such as CCA, RDA, and LFMM were conducted to identify the SNPs with potential adaptation to geographical and ecological variables. Manhattan plot revealed that many of these SNPs are significantly associated with these variables. The genetic changes accompanied by local and geographical adaptation were illustrated by skyline plot. These genetic changes occurred not under a molecular clock model of evolution, and probably under a positive selection pressure imposed in different geographical regions in which these plants grow.

Introduction

The genus *Avicennia* contains eight species of mangrove trees which mostly grow in intertidal zones of tropical and temperate regions. The mangroves are important to human due to their medical uses. Available medicinal literature indicates that different parts of these plants have been used for curing and treatment in various kind of diseases such as cancer, diabetes, malaria, rheumatism, asthma, small pox and ulcer (Hrudayanath et al. 2016). These species show variation and forms a complex group due to extensive geographical distribution and frequent inter-specific hybridization (Mori et al. 2015).

Several studies have been carried out to identify the putative adaptive SNPs in different plant species (see for example Abebe et al. 2015; Garot et al. 2021). These landscape genetic studies screen the genomes to identify differentiated regions (i.e. outlier loci) that are putatively under natural selection and test for associations between putative adaptive loci (e.g. SNPs) and environmental variables of the species habitat while accounting for neutral patterns that affect allelic frequencies such as genetic structure and demographic history. Therefore, these studies not only identify the candidate loci for adaptation but also identify the ecological selective pressures responsible for local adaptation (Garot et al. 2021).

Landscape genetic studies utilize different statistical and bioinformatics methods. For example, MDS (Multidimentional scaling), and PCA (Principal components analysis), have been used for the population divergence, while a combination of RDA (Redundancy analysis), and LFMM (Latent factor mixed model) have been used for identifying the adaptive genetic regions (see for example, Abebe et al. 2015;Garot et al. 2021). To our knowledge, there has been no report on putative local adaptive SNPs in The genus *Avicennia*, therefore, the present study was performed with the following tasks: 1- To Identify the nucleotides which can differentiate *Avicennia* species and geographical populations from each other, 2-

To reveal association between DNA sequences and geographical coordinates, and 3- To identify the SNPs which halve phylogenetic signal.

We used both PCA (Principal components analysis) and DAPC (Discriminate analysis of principal components analysis), which is suitable for SNP sequences, to identifying discriminating sequences. For association studies, we used CCA (Canonical correspondence analysis), and RDA (Redundancy analysis), followed by LFMM (Latent factor mixed model) analysis to test significance of nucleotide association with geographical and ecological variables. The phylogenetic signal of sequences was investigated by character mapping based on parsimony criterion.

Materials And Methods

In this study, we used published data on *ITS* sequences for a *Avicennia* species which are reported from different parts of the world in NCBI site in NCBI (Table 1).

Table 1
Voucher information and GenBank accession numbers of taxa sampled for the genus *Avicennia* based on ITS data.

Sp	Accession number
Avicennia alba	EF540977.1
Avicennia alba	AF365980.1
Avicennia alba	MH243937.1
Avicennia alba	MH243936.1
Avicennia alba	MH243935.1
Avicennia alba	MH243934.1
Avicennia alba	MG880036.1
Avicennia alba	MG880035.1
Avicennia alba	MG880034.1
Avicennia alba	MG880033.1
Avicennia alba	MG880032.1
Avicennia alba	MG880031.1
Avicennia alba	MG880030.1
Avicennia alba	MG880029.1
Avicennia alba	MG880028.1
Avicennia alba	EU528876.1
Avicennia alba	KX641594.1
Avicennia alba	KJ784551.1
Avicennia alba	KF848261.1
Avicennia bicolor	EF540988.1
Avicennia bicolor	EF540988.1
Avicennia bicolor	EF540987.1
Avicennia bicolor	EU352151.1
Avicennia bicolor	EU352150.1
Avicennia bicolor	EU352149.1
Avicennia bicolor	AF365977.1

Sp	Accession number
Avicennia bicolor	EU528877.1
Avicennia germinans	EF540990.1
Avicennia germinans	EF540985.1
Avicennia germinans	EF540984.1
Avicennia germinans	EF540983.1
Avicennia germinans	EF540982.1
Avicennia germinans	EF540981.1
Avicennia germinans	EF540980.1
Avicennia germinans	EU352146.1
Avicennia germinans	EU352147.1
Avicennia germinans	KX641596.1
Avicennia germinans	MG880047.1
Avicennia germinans	MG880046.1
Avicennia germinans	MG880045.1
Avicennia germinans	MG880041.1
Avicennia germinans	MG880040.1
Avicennia germinans	MG880039.1
Avicennia germinans	MG880038.1
Avicennia germinans	MG880037.1
Avicennia germinans	DQ469846.1
Avicennia germinans	DQ469845.1
Avicennia germinans	DQ469860.1
Avicennia germinans	DQ469859.1
Avicennia germinans	DQ469858.1
Avicennia germinans	DQ469857.1
Avicennia germinans	DQ469856.1
Avicennia germinans	DQ469855.1
Avicennia germinans	DQ469854.1

Sp	Accession number
Avicennia germinans	DQ469853.1
Avicennia germinans	DQ469852.1
Avicennia integra	KX641598.1
Avicennia officinalis	MH243949.1
Avicennia officinalis	MH243948.1
Avicennia officinalis	MH243947.1
Avicennia officinalis	MH243946.1
Avicennia officinalis	MH243945.1
Avicennia officinalis	MH243944.1
Avicennia officinalis	MH243943.1
Avicennia officinalis	MG880054.1
Avicennia officinalis	MG880053.1
Avicennia officinalis	MG880052.1
Avicennia officinalis	MG880051.1
Avicennia officinalis	MG880050.1
Avicennia officinalis	KX641597.1
Avicennia officinalis	KJ784553.1
Avicennia officinalis	KF848263.1
Avicennia rumphiana	KX641595.1
Avicennia schaueriana	EF540986.1
Avicennia schaueriana	DQ469862.1
Avicennia schaueriana	AB861412.1
Avicennia schaueriana	AB861385.1
Avicennia schaueriana	AB861382.1
Avicennia schaueriana	AB861365.1
Avicennia schaueriana	AB861357.1
Avicennia schaueriana	AB861354.1
Avicennia schaueriana	AB861345.1

Sp	Accession number
Avicennia schaueriana	AB861327.1
Avicennia schaueriana	AB861326.1
Avicennia schaueriana	AB861325.1
Avicennia schaueriana	AB861307.1
Avicennia schaueriana	AB861306.1
Avicennia schaueriana	AB861305.1
Avicennia schaueriana	AB861287.1
Avicennia schaueriana	AB861286.1
Avicennia schaueriana	AB861285.1
Avicennia schaueriana	AB861284.1
Avicennia schaueriana	AB861280.1
Avicennia schaueriana	AB861270.1
Avicennia schaueriana	AB861266.1
Avicennia schaueriana	AB861265.1
Avicennia schaueriana	AB861263.1
Avicennia schaueriana	AB861257.1
Avicennia schaueriana	AB861251.1
Avicennia schaueriana	AB861246.1
Avicennia schaueriana	AB861245.1
Avicennia schaueriana	AB861244.1
Avicennia schaueriana	AB861240.1
Avicennia schaueriana	AB861231.1
Avicennia schaueriana	AB861226.1
Avicennia schaueriana	AB861225.1
Avicennia schaueriana	AB861224.1
Avicennia schaueriana	AB861222.1
Avicennia schaueriana	AB861220.1
Avicennia marina	MF063712.1

Sp	Accession number
Avicennia marina	MF063711.1
Avicennia marina	MF063710.1
Avicennia marina	MF063709.1
Avicennia marina	MF063708.1
Avicennia marina	EF540978.1
Avicennia marina	AF477771.1
Avicennia marina	AF477770.1
Avicennia marina	MN883387.1
Avicennia marina	MN883386.1
Avicennia marina	MN883385.1
Avicennia marina	MN883384.1
Avicennia marina	MH243942.1
Avicennia marina	MH243941.1
Avicennia marina	MH243940.1
Avicennia marina	MH243939.1
Avicennia marina	MH243938.1
Avicennia marina	MG880049.1
Avicennia marina	MG880048.1
Avicennia marina	MK027295.1
Avicennia marina	EU528879.1
Avicennia marina	KM652500.1
Avicennia marina	KF848262.1
Avicennia marina	DQ469861.1
Avicennia marina subsp. *marina*	KX641593.1
Avicennia marina subsp. *eucalyptifolia*	KX641592.
Avicennia marina subsp. *australasica*	KX641591.
Avicennia marina subsp. *australasica*	AF365978.1

Data analyses

DNA sequences obtained were initially aligned by MUSCLE program implemented in MEGA 7 program and cured accordingly. These sequences were then used to construct Maximum likelihood phylogenetic tree (ML tree), by MEGA 7 program based on Kimura-Two parameters distance. The following analyses were performed to identify the SNPs which discriminate different *Avicennia* species and geographical regions and reveal association between SNPs and geographical coordinates of *Avicennia* species distribution. We should mention that these analytical approaches have different assumptions and may differ to some extent in their results. Therefore, comparing obtained results are important for drawing a solid conclusion.

Canonical correspondence analysis

In the first approach we used CCA (Canonical correspondence analysis), which is basically based on regression of the SNPs and ecological features. It uses an approach similar to principal components analysis (PCA), but is suited for discrete characteristics like SNPs (Podani 2000). In PCA we have a maximized variance of data, while CCA tries to maximize the association of data (SNPS), to geographical variables (Podani 2000). CCA was performed in PAST ver. 4. program (Hammer et al. 2012).

Redundency analysis (RDA)

RDA is a form of constrained ordination that suits for genomic data sets, where we are interested in understanding how the multivariate environmental factors shape the patterns of genomic composition across geographical areas. RDA is based on multivariate regression, and takes in to account a model of linear combinations of the environmental predictors that explain linear combinations of the SNPs. This method effectively identifies covering loci associated with the multivariate geographical variables (Legendre & Legendre 2012). RDA was performed in PAST ver. 4, program.

Latent Factor Mixed Model (LFMM)

LFMM is a Bayesian approach method for testing associations between loci and geographical gradients using latent factor mixed models. It performs a regression analysis in which the confounding variables are modeled with unobserved (latent) factors. The program estimates correlations between geographical and ecological variables and allelic frequencies, and simultaneously infers the background levels of population structure (Frichot et al. 2013, Frichot & Francois 2015). LFMM was performed by LFMM package in R. 4.2.

Phylogenetically important SNPs was determined by character mapping of 110 SNPs obtained based on parsimony criterion as performed in Mesquite 3.6 program. We performed Tajima's D test (Tajima 1989), to reveal if *Avicennia* species DNA sequences evolved randomly (“neutrally”), or under a non-random process, including directional or balancing selection, demographic expansion or contraction. Moreover, we also carried out the molecular clock test, to show if SNP changes occurred in accordance with a uniform clock rate model of evolution during *Avicennia* genus speciation events. These tests were

performed by MEGA 7 program. The skyline analysis was used to study population size changes in different geographical regions as performed in R-package 4.2.

Results

Geographical grouping of Avicenna species

We obtained DNA sequences of 540 nucleotides length after alignment and curation, of which, 116 nucleotides where polymorphic among the studied *Avicennia* genotypes. The ML phylogenetic tree based on these sequences (Fig. 1), revealed that species from some of geographical regions can be differentiated as they comprise a specific clade on ML tree. This is almost entirely true for taxa studied from Brazil, and also to a great extent for genotypes of Bangladesh and India, as well as Saudi-Arabia and Egypt. However, the studied species and genotypes from the other geographical areas like Mexico, Costa-Rica, and Panama, were placed inter-mixed but in a distinct clade. These results suggest that we may find some private sequences for these particular geographical regions.

Dentrented correspondence analysis (DCA) of the sequences (Fig. 2), revealed the most variable ITS nucleotides among *Avicennia* genotypes studied. These nucleotides are scattered throughout this genetic region as they do not cluster to each other. Therefore, different segments of ITS sequences can differentiate *Avicennia* taxa. Some of these nucleotides shoe correlation to each other as they are placed closer to each other and in a particular direction of DCA plot. For example, nucleotides number 21, 29, 49, and 109, all are placed closer to each other on the left side of DCA plot. The same holds true for nucleotides number 37, 97, 102, 104, and 104, which are placed close to one another but in an opposite direction to the previously mentioned nucleotides. In total, we can say that out of 110 polymorphic nucleotides, about 25–30 are showing a higher degree of discriminating power.

DAPC analysis (Fig. 3), revealed that individuals studied from Bangladesh and India, show some degree of genetic affinity. The same holds true for the samples studied from Brazil and the other Latin America countries. DAPC analysis also revealed that the first three axes comprise about 80% of total variance and based on these axes, several SNPs were identified that can differentiate *Avicennia* taxa. Four nucleotides namely, 48, 88, 89, and 90, showed the highest degree of discriminating power.

CCA and RDA analyses produced similar results. Moreover, RDA produced significant (0.05) association between SNPs and geographical as well as ecological variables (Fig. 4). Therefore, the RDA results are presented here.

RDA plot (Fig. 4), revealed that some of these discriminating nucleotides are associated with longitudinal distribution of *Avicennia* species, while some others are associated with latitudinal distribution. This plot also shows that some of these nucleotides are potentially associated with a particular geographical region, like South-America, or East Asia. For example, nucleotides number 52, 55, 67, 72, and 76, are associated with longitudinal distribution of *Avicennia* species. The other nucleotides which are associated with latitude and a particular geographical area are also indicated in this Figure.

FLMM (Fig. 5) analysis revealed a highly significant association between most of the discriminating nucleotides suggested by CCA and RDA analyses. The Manhattan plot of LFMM revealed that most of these nucleotides have -log 10 p value of 1 and higher, which are highly significant and correlated with *Avicennia* geographical regions.

Detailed inspection of these nucleotides by parsimony analysis revealed that they are mostly private or restricted to a particular country. For example, nucleotide number 80, which is a C nucleotide for all the studied species, is different only in Madagascar samples with A Nucleotide. Similarly, the nucleotides positions 24 and 27 are C for all geographical regions except for Brazil which is a T nucleotide.

Tajimas' D produced a positive value (D = 0.32), which indicates a positive selection over the sequences. Similarly, when data were subjected to Neighbor-Net network analysis (Fig. 6), almost two major groups were formed, which supports Tajimas, D, and the presence of a positive selective force over sequences in *Avicennia* species.

The molecular clock test showed that SNP changes within the genes *Avicennia* did not occurred under uniform rate of evolution and different phylogenetic clades differed in their genetic changes. This results also agree with the result of skyline plot (Fig. 7), showing a deep and a sudden change in SNP substitution and population size change in *Avicennia* species in different geographical regions.

Discussion

Speciation within the genus *Avicennia* is complex. Recently, (unpublished data), we provided DNA barcode of ITS region, in *Avicennia* species which illustrate genetic differentiation between taxa growing in different geographical regions.

In spite of many studies performed on the phylogenetic and genetic diversity of *Avicennia* genus, there has been no report on geographical association of DNA sequence in this genus. The present study revealed that DNA nucleotides of ITS region can efficiently differentiate geographical taxa in *Avicennia* genus. Moreover, some of these sequences may be significantly associated with geographical distribution of these species.

Tajima's test of these sequences produced a positive Tajima's D, which indicates a balanced selection related to a speciation events (Tamura and Nei 1993). We also observed almost a continuous and gradual nucleotide substitution for some of the species growing in some parts of the world, but with a sudden deep change in DNA sequences in some other geographical regions. This was supported by Neighbor Net plot.

Different approaches used to identify the nucleotides associated with geographical and ecological variables, were all in agreement, and LFMM Manhattan plot showed significant association between some of the SNNPs and ecological as well as geographical variables. The same results were obtained in spite of the fact that

the analytical approaches used differ in their computation approaches. CCA and RDA methods are based on linear association (regression), while LFMM method is a Bayesian-Model approach (Podani 2000; Frichot and Francois 2015). It seems therefore, using different approaches may improve understanding of associated SNPs with geographical and ecological variables and such a combined data evaluation, give insights into contemporary evolutionary processes, and may explain how environmental factors influence selective and neutral genomic diversity within and among related species or different geographical populations within a single species (Segovia et al. 2020).

Presence of heterogenous environmental conditions bring about changes in the genetic diversity of plant species which in turn results in local adaptations (Segovia et al. 2020). Therefore, the studies concerned with the genetic basis of local adaptation and identifying adaptive genetic loci or SNPs can improve the knowledge of the genetic mechanism of local adaptation and probably species diversification within a genus (Zhang et al. 2019). Moreover, these studies try to answer two major questions: 1- which environmental variables play key role in the adaptive genetic divergence of a species or different species within a particular genus and shape its landscape genetic structure, and, 2-which genes or loci on the genome undergo adaptive genetic differentiation (Li. et al. 2017; Zhang et al. 2019).

The present study revealed that both longitudinal as well as latitudinal distribution of *Avicennia* species as well as ecological variables like temperature, and rainfall have selective pressure on the studied SNPs and play role in genetic changes within this genus.

In some plant species, we may encounter the influence of one of these geographical variables on sequence adaptation. For example, Ingvarsson et al. (2006), characterized patterns of DNA sequence variation at the putative candidate gene phyB2 in 4 populations of European aspen (*Populus tremula*) and scored single-nucleotide polymorphisms in an additional 12 populations collected along a latitudinal gradient in Sweden. They utilized a sliding-window scan of phyB2 and identified six putative regions with enhanced population differentiation and four SNPs showed significant clonal variation. Therefore, they suggested that the cline variation at individual SNPs is an adaptive response in phyB2 to local photoperiodic conditions. It has been suggesting that divergent selection enhances the levels of genetic differentiation not only for the sites that are the direct target of selection but also for neutral sites in the vicinity of the site(s) under selection (Charlesworth et al. 1997; Nordborg and Innan 2003).

In conclusion, the present study provide data on DNA sequences changes in association with geographical and ecological variables in the genus *Avicennia* and suggest that these variables play role in causing genetic changes within this genus. Some of these SNPs were significantly associated with these geographical and ecological variables.

Declarations

Funding

None

Conflicts of interest

All authors have no conflict of interest

Availability of data and material

All data are available from the corresponding author on reasonable request.

Code availability

Not applicable

Author contribution statement

Masood Sheidai: conceptualization of the project and corresponding author, Farrokh Ghahremaninejad: conceptualization of the project and data collection, Afshin Danehkar: conceptualization of the project and plant collection, Fahimeh Koohdar: conceptualization of the project and lab work, Laleh malekmohammadi: data collection and lab work

Ethics approval

Not applicable

Consent to participate

Not applicable

Consent for publication

Not applicable

References

1. Abebe T D, Naz AA, Léon J (2015) Landscape genomics reveal signatures of local adaptation in barley (*Hordeum vulgare* L.)
2. Garot E, Dussert S, Domergue F, Joët T, Fock-Bastide I, Combes MC, Lashermes P (2021) Multi-Approach Analysis Reveals Local Adaptation in a Widespread Forest Tree of Reunion Island. Plant Cell Physiol. 62 (2): 280–
3. Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet Res. 70: 155–174.
4. Frichot E, Francois O (2015) LEA: An R package for landscape and ecological association studies. In. press.

5. Frichot E., Schoville S.D., Bouchard G.& FrancoisO. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 30 (7): 1687–1699.
6. Hrudayanath T, Dibyajyoti S, Swagat KD (2016) The genus *Avicennia*, a pioneer group of dominant mangrove plant species with potential medicinal values. Front Life Sci. 9: 267–291.
7. Ingvarsson P K, Garcı´a MV, David 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.
8. Legendre P, Legendre L (2012) Numerical ecology. 3 editions. Oxford, UK.
9. Li X, Duke NC, Yang Y, Huang L, Zhu Y, Zhang Z, Zhou R, Zhong C, Huang Y, Shi S (2016) Re-Evaluation of Phylogenetic Relationships among Species of the Mangrove Genus *Avicennia* from Indo-West Pacific Based on Multilocus Analyses. PLoS ONE. 11(10): e0164453.
10. Li Y, Zhang XX, Mao RL, Yang J, Miao CY, Li Z, Qiu XY (2017) Ten years of landscape genomics: challenges and opportunities. Front Plant Sci. 8: 2136.
11. Mori GM, Zucchi MI, Sampaio I, Souza AP (2015) Species distribution and introgressive hybridization of two *Avicennia* species from the Western Hemisphere unveiled by phylogeographic patterns. BMC Evol Biol. 15: 61.
12. Nordborg, M, Innan H (2003) The genealogy of sequences containing multiple sites subject to strong selection in a subdivided population. Genetics. 163: 1201–1213.
13. Podani J (2000) Introduction to the Exploration of Multivariate Data Backhuyes, Leiden.
14. Segovia NI, González–Wevar CA, Haye PA (2020) Signatures of local adaptation in the spatial genetic structure of the ascidian Pyura chilensis along the southeast Pacific coast. Sci Rep. 10: 14098.
15. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123: 595–595.
16. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10: 512–526.
17. Zhang X X, Liu B G, Li Y, Liu Y, He YX, Qian Z H, Li J X (2019) Landscape genetics reveals that adaptive genetic divergence in *Pinus bungeana* (Pinaceae) is driven by environmental variables relating to ecological habitats. BMC Evol Biol.19:160.

Figures

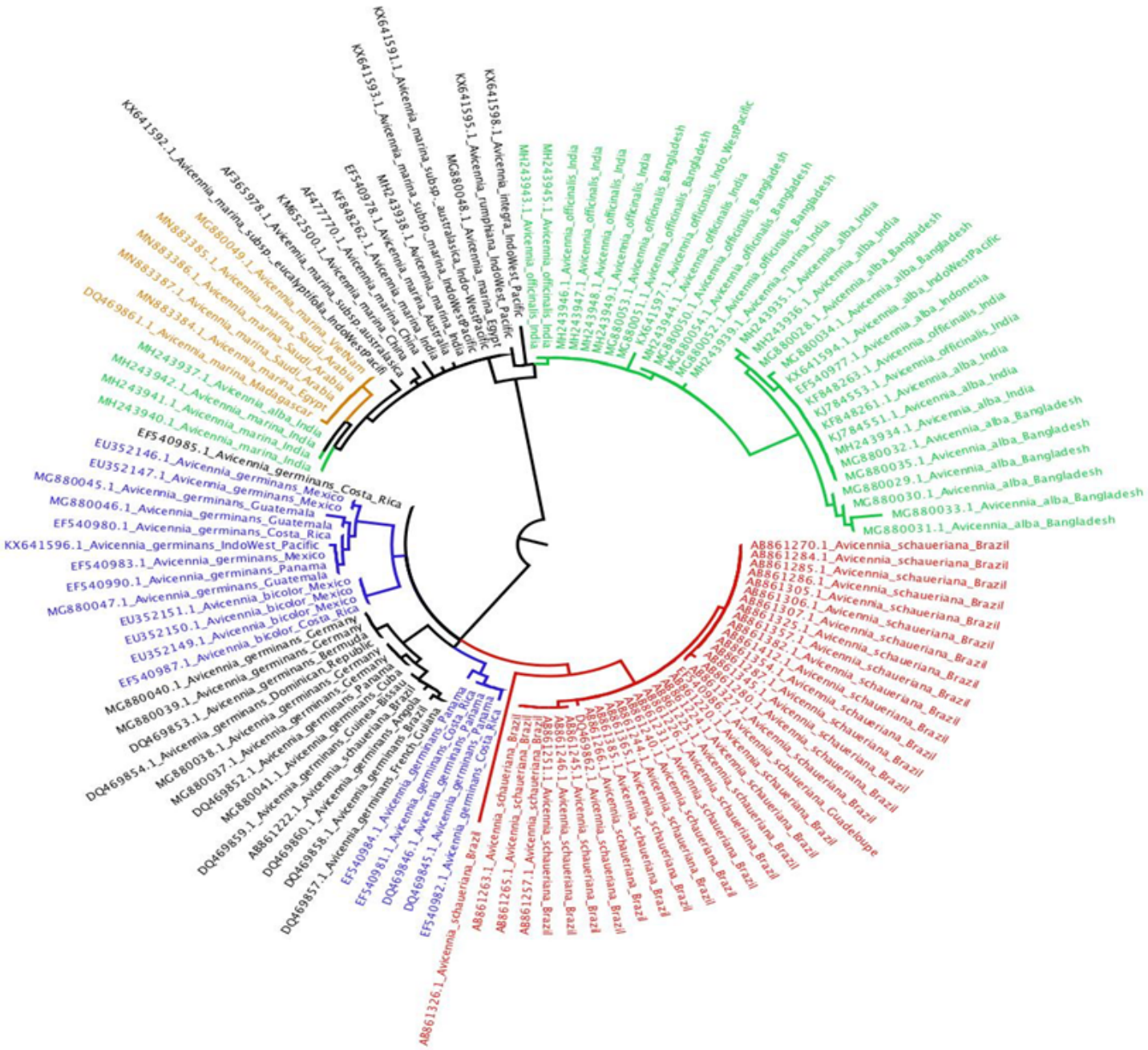

Figure 1

ML phylogenetic tree of *Avicennia* genotypes based on their geographical regions.

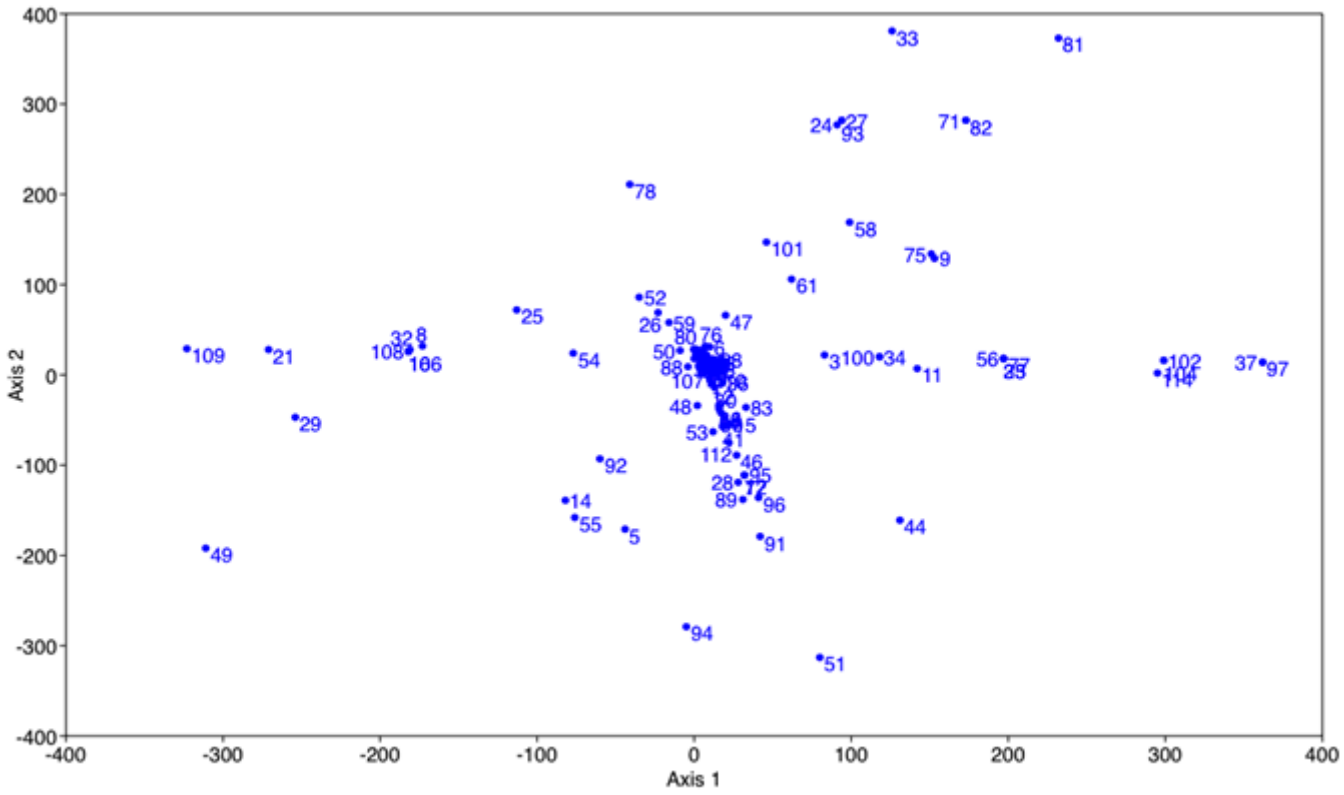

Figure 2

DCA plot of ITS sequences showing nucleotides with a higher degree of discriminating power among *Avicennia* genotypes.

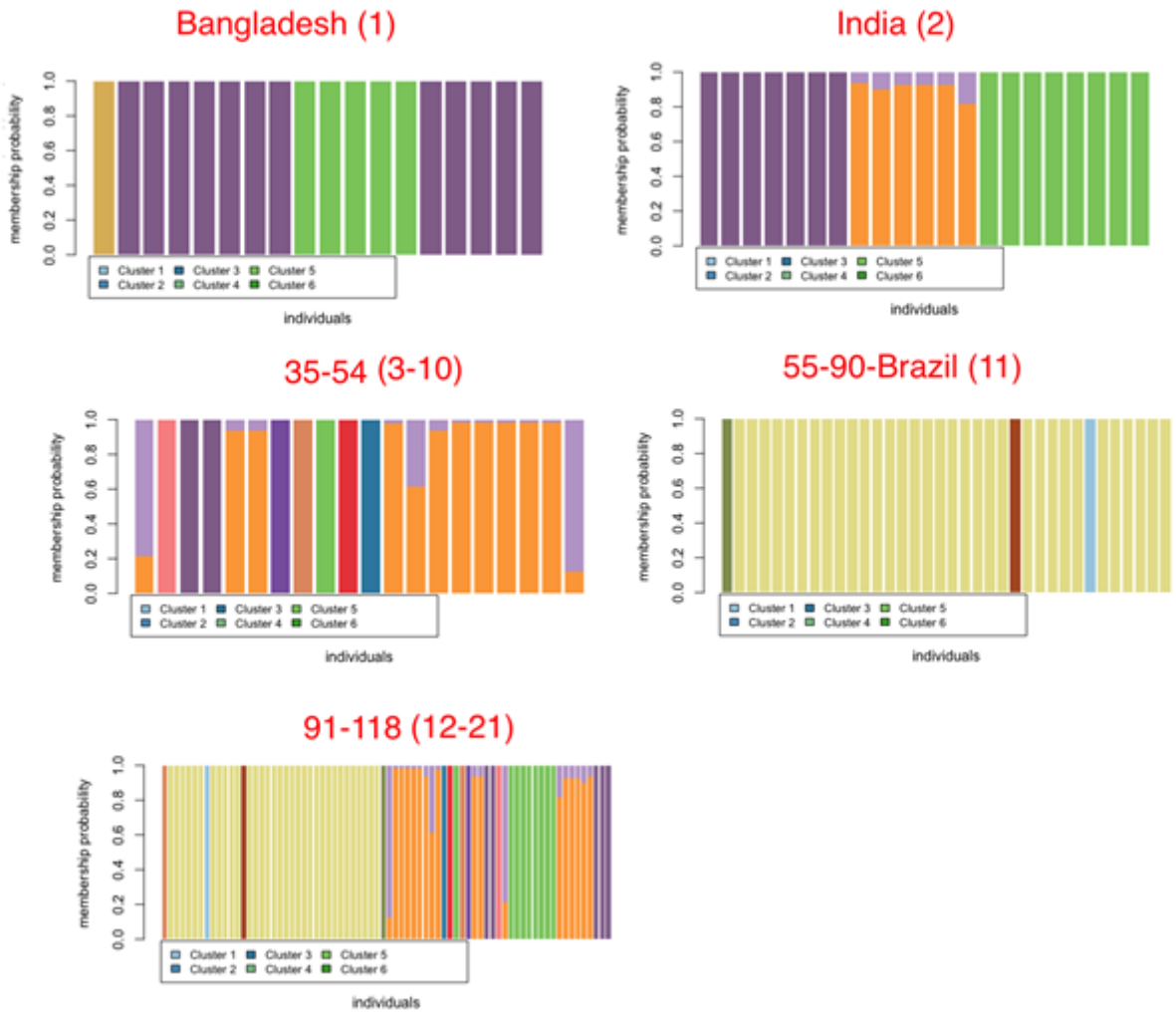

Figure 3

DAPC genotype assignment of *Avicennia* species based on geographical regions.

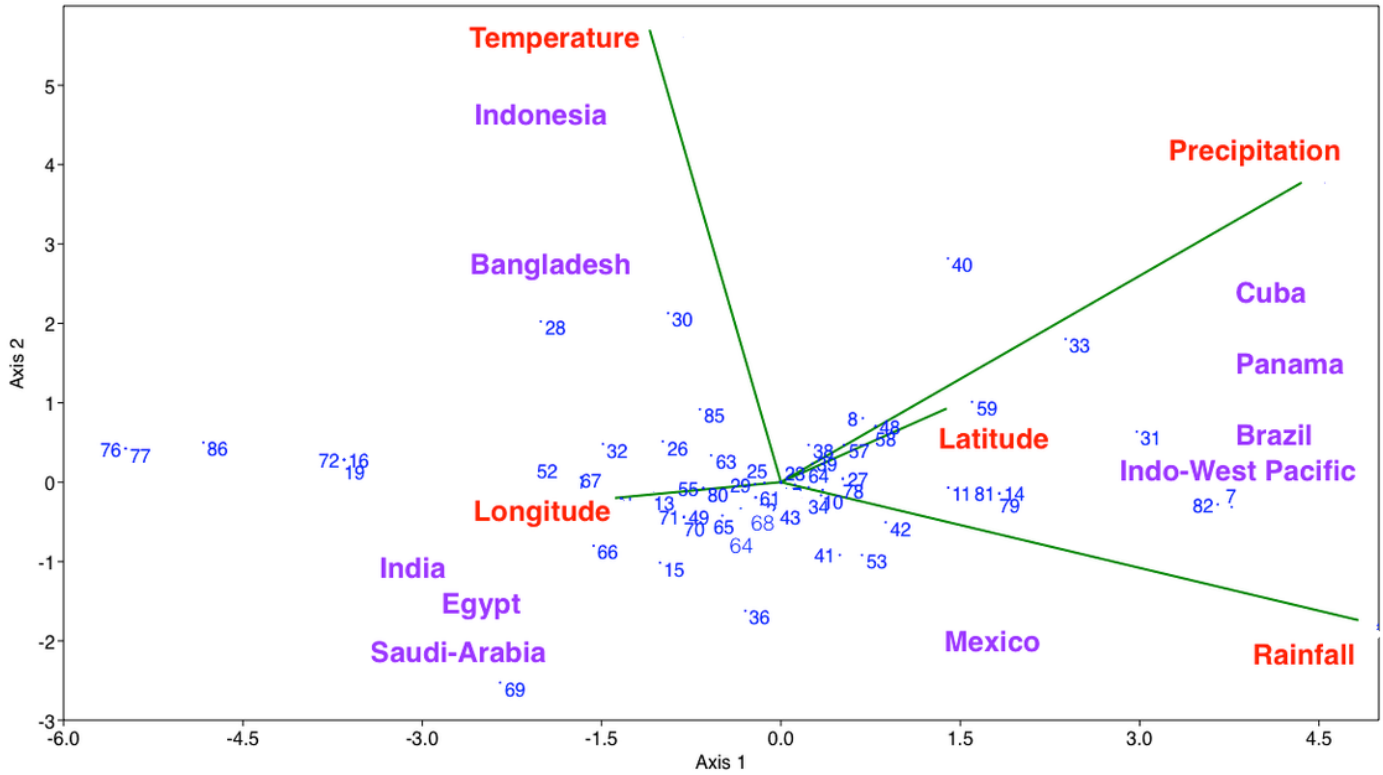

Figure 4

RDA plot of *Avicennia* species showing SNPs with are potentially associated with different geographical regions.

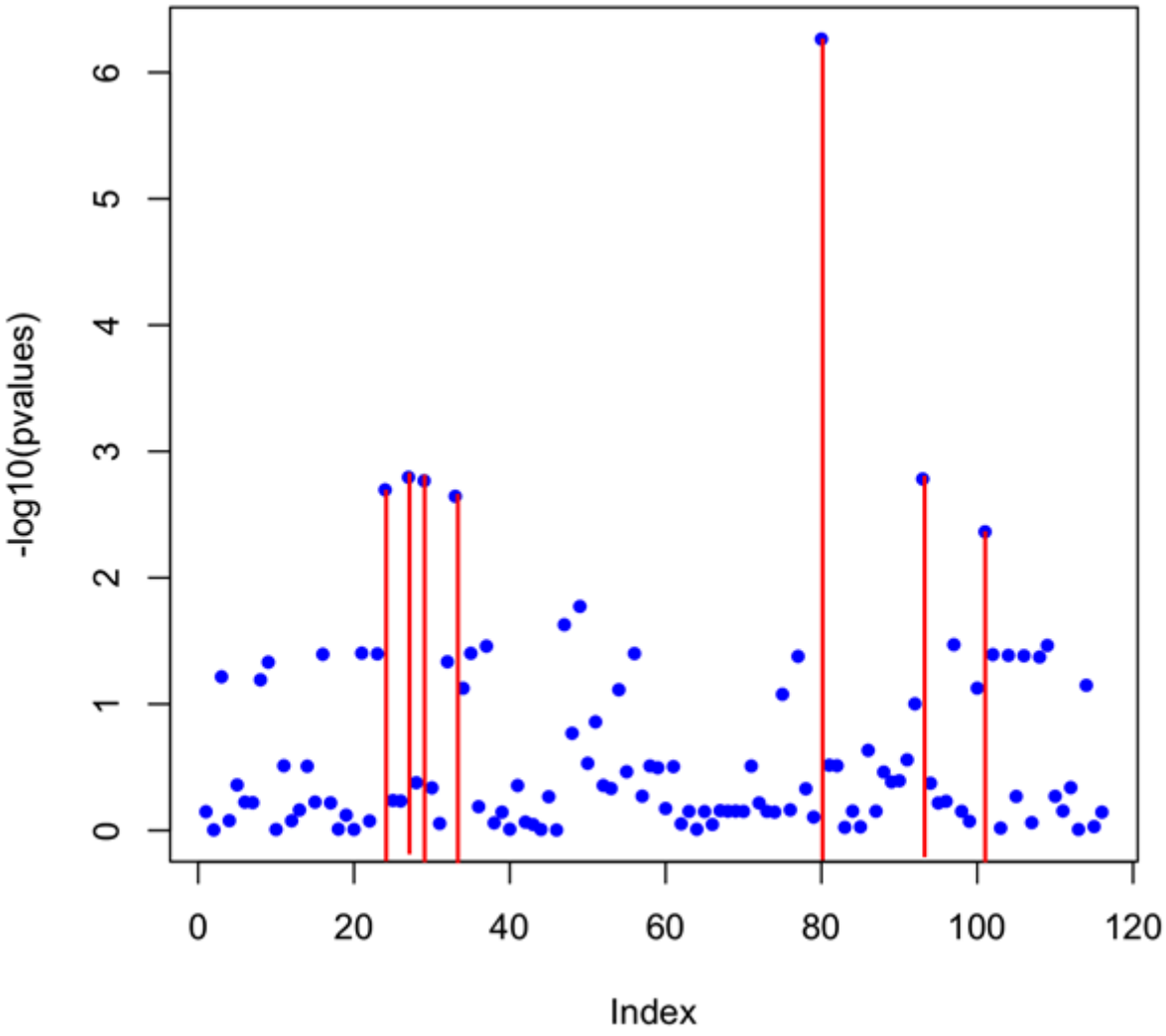

Figure 5

Manhattan plot of LFMM analysis showing significantly associated SNPs with geographical variables.

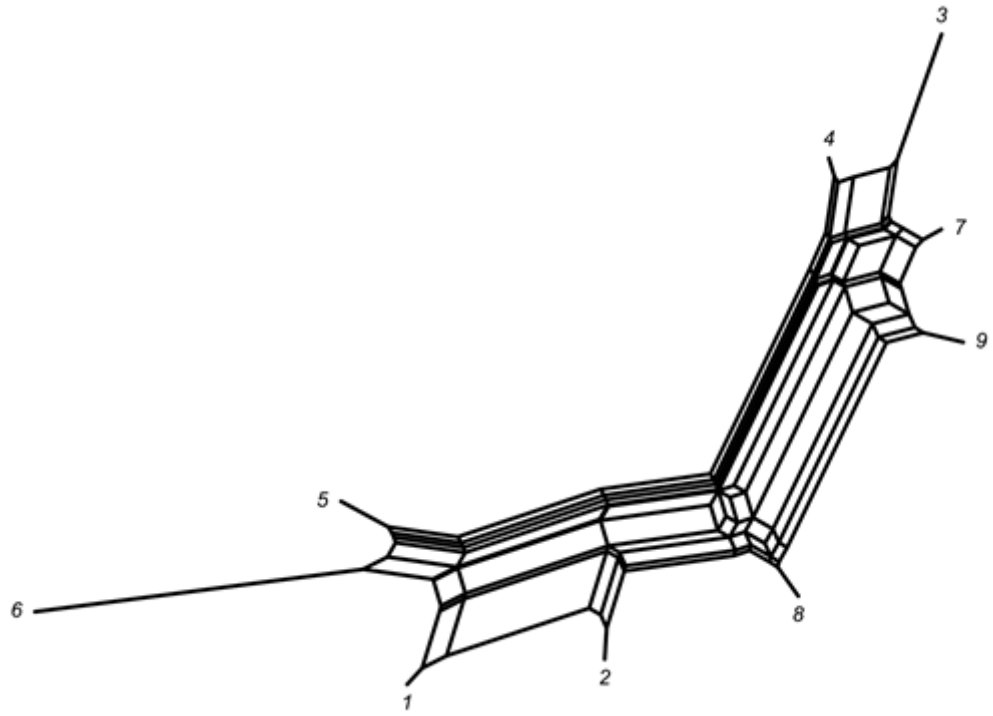

Figure 6

NeighborNet diagram of *Avicennia* species showing two major groups formed in accord to different geographical regions. (Species 1, 2, 5, 6, and 7 are the samples studied from Asia and Arabian countries, while the other sample are from South-America).

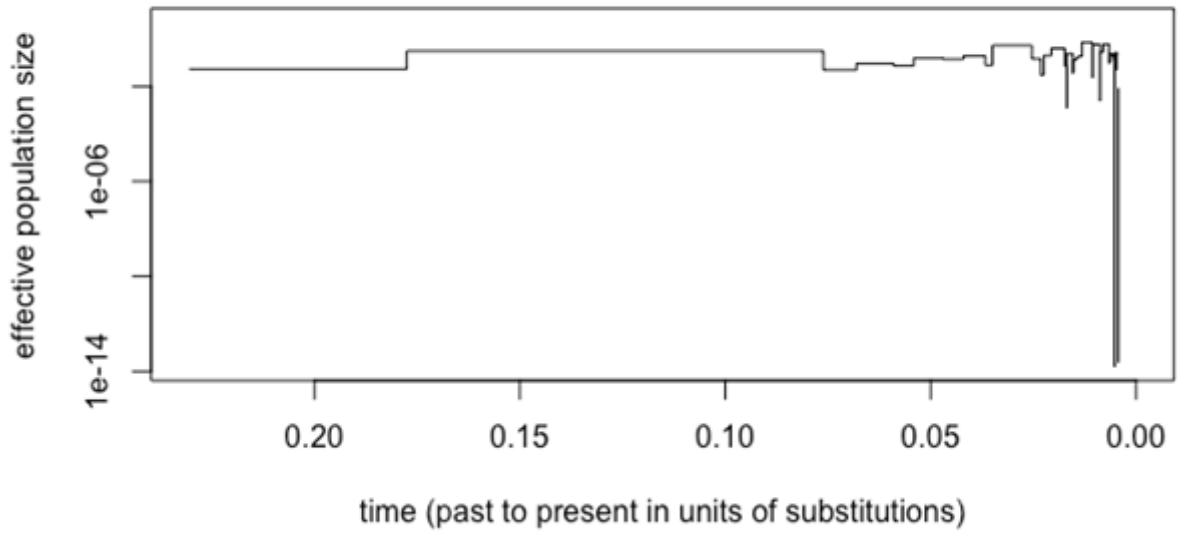

Figure 7

Skyline plot of *Avicennia*. species showing abrupt sudden change in some of the species due to geographical regions they grow in.