Agriculture, Ecosystems and Environment 139 (2010) 51–58

Contents lists available at ScienceDirect

Agriculture, Ecosystems and Environment

journal homepage: www.elsevier.com/locate/agee

Quantifying the global environmental niche of an underutilised tropical fruit tree (*Tamarindus indica*) using herbarium records

C. Bowe*, N. Haq

Centre for Underutilised Crops, Environment Division, School of Civil Engineering and the Environment, University of Southampton, Southampton SO17 1BJ, UK

ARTICLE INFO

Article history:
Received 15 May 2009
Received in revised form 10 June 2010
Accepted 28 June 2010
Available online 23 July 2010

Keywords:
Herbarium data
Data quality
Niche evolution
Climate adaptation
Spatial non-stationarity
Underutilised tropical fruit tree species

ABSTRACT

The importance of identifying the environmental adaptation of underutilised species such as tamarind (*Tamarindus indica* L.) and their ecogeographic distribution is widely recognised. However the lack of physiological or quantitative yield/growth data does not permit the use of more traditional methods of crop modelling. In this study a representative dataset of tamarind's global distribution appropriate for modelling the species distribution is produced mainly from herbarium records; while minimising the effects of error and bias inherent in such data. The exploratory analysis of the clean dataset showed that both at the bioregional and sub-regional scale tamarind is experiencing varying conditions in different regions. This indicates the existence of spatial non-stationarity, which due to the broad distribution and semi-domesticated nature of the species, could be caused by evolution of the niche. Spatial niche variation in tamarind indicates that its range may have expanded from its naturalised area in Africa and India eastward into South East Asia and Latin America, progressively moving into regions with a less pronounced dry period and wetter conditions. Difference in environmental niche between East African and West African populations may be caused by the Rift Valley preventing gene flow between the two regions of the continent. Such information can be related to genetic variation and structuring within such species, providing information useful for conservation and selection of plant material for adaptation to future climate. The implication of spatial non-stationarity and spatial variation within the niche, when considering if global or regional/multi-scale models should be used to predict the world distribution of such species under current and future climate scenarios, are discussed.

1. Introduction

The importance of identifying environmental adaptation of underutilised species and their eco-geographic distribution has been widely recognised (Padulosi et al., 1999; Hall and O'Brien, 2002; Williams and Haq, 2002). Many underutilised species are adapted to harsh conditions, provide biodiversity and buffer against risk in a changing environment (Williams and Haq, 2003; Padulosi et al., 2003; Dawson et al., 2007). Consequently information regarding their environmental adaptation and ecogeographic distribution will be important in selecting crops for future utilisation.

Underutilised fruit trees in particular have the potential to diversify the farming system (Dawson et al., 2009), contribute to the economic development of poor communities and to alleviate hidden hunger (Jamnadass et al., 2009). The tamarind (*Tamarindus indica* L.) for example is a high value, multipurpose crop suited to diverse multifunctional rain-fed farming systems (El-Siddig et al., 2006). It can be grown on marginal land, is drought tolerant and provides a range of environmental services. Tamarind is potentially an ideal crop for resource-poor farmers across the tropics and subtropics but has received little research attention and is not a major contributor to global markets. However it is well known for its use as a food ingredient across the world and is very important for food security and livelihood creation for the rural poor; particularly for women who often play a major role in its collection and processing (El-Siddig et al., 2006). The main product is the fruit pulp; in addition the leaves, flower and bark also have numerous nutritional, industrial and medicinal uses (El-Siddig et al., 2006). The seeds contain phenolic compounds with antioxidant properties (Sudjaroen et al., 2005). Within Asia and the Americas, it is sold at local, national and international level and in Africa many countries have prioritised tamarind for conservation based on its value and utilisation (El-Siddig et al., 2006).

Tamarind has a life span of ca.150 years. Once fruiting begins at 10–12 years, the productive life of the tree can persist for 50–70 years. Tamarind is insect pollinated and highly outcrossing (with some level of selfing). It has broad-scale dispersal methods (animals and man) and its large hard seeds stay viable for months (El-Siddig et al., 2006). Tamarind is intimately related to people and habitation (El-Siddig et al., 1999; Muok et al., 2000; Misra and Dash, 2000;

---

* Corresponding author. Tel.: +44 02380 594653; fax: +44 023 8067 7519.
E-mail addresses: c.bowe@soton.ac.uk, cb13@soton.ac.uk (C. Bowe).

0167-8809/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.agee.2010.06.016

Codjia et al., 2003; Ahmed and Rahman, 2004) and is referred to as an incipient or semi-domesticate. It has a long and complicated history of natural and human expansion of its range, introduction and selection, in part due to its long term utilisation. It therefore has a broad geographical distribution and is adapted to a wide range of ecological conditions (El-Siddig et al., 2006).

Although recent research has focused on the genetic and phenotypic variation of underutilised fruit tree species (Harris et al., 1997; Dawson and Powell, 1999; Diallo et al., 2007; Jamnadass et al., 2009), little information is available on their ecological requirements, adaptive traits, adaptability and potential response to changing climates. Analysis of the species environmental niche and modelling the species' response and distribution under current and future climate scenarios should allow us to increase our knowledge in this regard. However little yield, growth or ecological data exists which would allow traditional methods of crop modelling such as yield-based empirical models (Drummond et al., 2000; Louw and Scholes, 2006) or process-based production models to be used (Stape et al., 2004; Park et al., 2005).

However, a large and increasingly accessible data source exists in the form of plant species locations or 'presence data' from herbarium records (Chapman, 1999; Graham et al., 2004). Collections of plant and animal specimens have been in existence for centuries in herbaria and museums. It is estimated that over 3 billion plant and animal records exist (Chapman, 1999), each with an associated 'collection event', describing the time and place where the specimen was found (Graham et al., 2004). These records span a substantial temporal and spatial extent and they are unique in that they cannot be recollected or obtained from any other source (Chapman, 1999). Several initiatives such as the Global Biodiversity Information Facility have used innovative information technology to connect multiple data providers and users within a single Internet site. Although traditionally used in taxonomic studies (Chapman, 1999; Williams et al., 2002), such data is increasingly being used in bio-geographical analysis (Loiselle et al., 2003; Van Zonneveld et al., 2009). The use of such data, when combined with environmental information, allows the quantitative modelling and analysis of species niche and potential distribution.

The aim of this study is to produce a representative dataset for the underutilised species tamarind, appropriate for analysis and modelling of species distribution and at a global scale; while minimising the effects of error and bias inherent in such a data source (Chapman, 2004). As larger extents are modelled, it is highly likely that heterogeneity in the predictor variables increases and that areas are included where species respond to habitats in different ways (Osborne and Suarez-Seoane, 2002; Estrada-Pena et al., 2006). This variation in relationships across space is known as spatial non-stationarity. Genetic variation in adaptive traits are likely to exist for plant species that cover large or environmentally heterogeneous distribution areas (Lillesø et al., 2001) and are subject to high levels of human introduction and selection (Peterson and Holt, 2003; Miller and Knouft, 2006) leading to niche variation. This will further increase the likely hood of spatial non-stationarity. We hypothesise that due to tamarinds' broad distribution and semi-domesticated nature, it is likely to be experiencing spatial non-stationarity across its global distribution. This is important as it would allow us identify if global or regional models are more suitable for predicting the distribution of widely distributed semi-domesticated underutilised tree species under current and future climate scenarios.

Analysis of the tamarind distribution dataset based on geographic region allows further knowledge about the species' niche and ecological adaptation to be derived. This paper will highlight the procedures that should be considered before using *ad hoc* sources such herbarium data for species distribution modelling. It is hoped that this paper will encourage users of such data to consider these important issues.

2. Methods

2.1. Environmental data

A set of meaningful ecological variables suitable for the analysis of the global distribution of underutilised tree species were collated and derived (Table 2). Climate data was obtained from Worldclim (Hijmans et al., 2004, 2005) and the Climate Research Unit (New et al., 2000). Quarterly variables for Relative Humidity and Sunshine were derived (New et al., 2000). Soil datasets with continuous values were produced by combining Dominant Soil Type and Topsoil Texture datasets from the TERRSTAT database (FAO, 2002a) with the ISRIC WISE Derived Soil Parameter Data Set (ver2.2) (Batjes, 2002). It should be noted that this dataset only considers the dominant soil type within each grid cell capturing only those characteristics that occupy 30–60% of the total grid area (Nachtergaele pers. comm.). A computer programme was written in $C^{\#}$ to calculate a Monthly Water Balance and produce Moisture Availability Index datasets, based on the method used in Global Agro-Ecological Zones project (FAO/IIASA, 2000). Reference Evapotranspiration (New et al., 2000; IMWI, 2005), 'Maximum Available Soil Moisture' (FAO, 2002b) and Monthly Rainfall Datasets (Hijmans et al., 2004) were used as input data for the programme. Actual Evapotranspiration (Eta) was calculated according to FAO (1979). The Monthly Reference Evapotranspiration (Eto) dataset was reclassified based on figures from the "Soil Water Depletion Fraction (p) for Crop Groups and Maximum Evaoptranspiration (ETm)" (FAO, 1979). Tamarind was assumed to be within group three due to being most commonly found growing in the same agro-ecological conditions as other crops within this group (FAO, 1979).

2.2. Species data

2.2.1. Collection and recording of species distribution information

The majority of the species records were collected from digitised herbarium records, available through biodiversity networks (Table 1). Records were also acquired from herbarium sheets from the collection at the Royal Botanical Gardens, Kew and 'mother tree' locations reported by Chanda (2001). In addition, tamarind locations were recorded using a Global Positioning System (GPS) during a field survey in India between June and September 2003.

2.2.2. Georeferencing species data and accounting for error and bias

Based on the text description of the herbarium data collated for this study all species records were geo-referenced. Geographical co-ordinates (latitude, longitude) were assigned to each location (Polhill, 1988; Pope and Pope, 1998; GNS-National Geospatial-Intelligence Agency, 2005) using the geodetic coordinate system 'GEOGRAPHIC LATLONG' and the datum WGS84.

The use of herbarium records in bioregional studies and specifically species distribution modelling has highlighted issues of data quality. Herbarium collections were not established for the purposes of modelling consequently *ad hoc* collection; non-systematic sampling and uneven sampling effort are common (Chapman, 1999; Stockwell and Peterson, 2001; Chapman, 2004; Wieczorek et al., 2004). Geographic co-ordinates are seldom given on herbarium specimen labels and collection localities have typically been recorded as textual descriptions (Chapman, 1999, 2004; Wieczorek et al., 2004). Adding the geographic co-ordinates after the collecting event can also produce errors relating to both accuracy and precision which, may lead to bias (Chapman, 1999, 2004; Wieczorek et al., 2004). The data set may have a temporal bias due to species no longer being present at a historical collection site (Wieczorek et al., 2004; Williams et al., 2002). Therefore rather than the sample population being representative of the species in environmental

Table 1
Digital herbarium and flora sources for tamarind occurrence records used in this study.

The Herbarium of the Institute of Ecology, A.C., Mexico	XAL
The Herbarium of the National Institute of Biodiversity of Costa Rica	INB
New York Botanical Garden	NY
Missouri Botanical Garden	MO
Herbarium of the Yucatan Scientific Research Centre, Mexico	CICY
Native Trees and Shrubs for Restoration and Reforestation of Mexico	IE – DF, UNAM
National Vegetable Germplasm Bank, Mexico	BANGEV, UACH
The Herbarium of the University of Texas, Austin, USA	TEX
Trees from the Yucatan Peninsula, flora from Tehuantepec, Oaxaca and Asteraceae family in Mexico	IBUNAM
The useful flora from two native communities of the Valley of Tehuacan-Cuicatlan	FES –I, UNAM
University of Florida Herbarium	FLAS
Western Australia Herbarium	PERTH
Australian National Botanic Gardens	CBG
Australian National Herbarium, Australia	CANB
Royal Botanic Gardens, Australia	MEL
State Herbarium of South Australia	AD
Western Australian Herbarium	PERTH
Flora Zambesiaca, Royal Botanical Gardens, Kew, UK	K
Universität Ulm	ULM
Cameroon National Herbarium, Yaounde	YA
Herbarium of the University of Aarhus	AAU
Oxford University Herbaria	OXF
Oxford Caribbean Specimens	OXF
South East Asia Botanical Collection Information Network	SEABCIN

or geographical space, it is instead dominated by the patterns at sampled sites (Williams et al., 2002; Barry and Elith, 2006).

However a good understanding of potential sources of error can lead to active quality control and managed improvement in overall data quality (Burrough and McDonnell, 1998). To identify and attempt to correct or remove erroneous location records, a number of methods were used to identify outliers in geographic and environmental space. Outliers in geographic space were identified by visual analysis. The distribution points and administrative unit datasets (ESRI, 2000b; GNS-National Geospatial-Intelligence Agency, 2005) in ARC View 3.2 (ESRI, 2000a) were overlaid and those points which appeared to be outside the geographic range of the species or those that appeared to fall in the sea were flagged and checked. The 'assign data by location' function (ESRI, 2000a) was used to join the species point distribution dataset to both gazetteer and administrative unit datasets (ESRI, 2000b; GNS-National Geospatial-Intelligence Agency, 2005) based on their underlying spatial relationships. The new dataset created by this 'join' was then analysed to ensure that the species location description, (i.e. place name, district, state) matched with the description assigned from gazetteer and administrative unit datasets.

To identify outliers in environmental space values from the underlying environmental grids were written to the corresponding point in the species location table, creating a dataset of the environmental conditions at each of the presence locations. This was done using the programme 'Grid Sampler' (Zerger, 2004). A statistical reverse jack-knifing procedure (Barnett and Lewis, 1978) and a cumulative frequency graph analysis were carried out for each variable to identify outliers in univariate space (Chapman, 1999) and a Principle Component Analysis (McGarigal et al., 2000) with all variables used to identify outlier in multivariate space. If the critical value was greater than a threshold value, within the reverse jack-knifing procedure the record is regarded as an outlier and flagged as suspect (Chapman, 1999). Specimen records marked on the cumulative frequency graph which created an elongation of the sigma curve were flagged. For the PCA analysis outliers (points distant from the main cluster) were identified based on visual analysis of component score plots and flagged. All flagged records had their geo-references checked. If correction was not possible, the record was removed from the dataset and further analysis.

The species dataset was converted to 10 min resolution (one record per grid square), the same resolution as the environment datasets. This helped to account for some of the sampling bias associated with the dataset while additionally addressing geo-referencing precision (georeferencing of records need only be precise to the scale of minutes). To account for temporal bias only records from 1950 and later were used. This coincides with the periods used to calculate the averages for the Worldclim climate dataset. The 1950 cut-off date allowed over 50% of the occurrence records to be kept maintaining a workable sample size.

2.3. *Data exploration*

Once all herbaria records were checked and suspect records were removed, a statistical summary for each environmental variable was calculated. Each presence record was also categorised based on its biogeographic region (Udvardy, 1975) using shape file (WCMC, 2001) and United Nation macro-geographic (continental) sub-regions (United Nations Statistics Division, 2008) in Arc view 3.2 (ESRI, 2000b). A principal component analysis was run for all 359 data points and for a subset dataset from tamarinds' native range (Sub-Saharan Africa, India and Sri Lanka) to investigate relationships between biogeographic region/sub-regions and tamarind's niche requirements.

3. Results

3.1. *Final tamarind occurrence dataset*

702 presence points remained after the removal of erroneous records. This was reduced to 543 points with removal of records collected before 1950. Once converted to 10 min resolution this was reduced to 443 points, 359 of which occurred at locations where values were available for all environmental variables (Fig. 1).

3.2. *Data exploration of tamarind occurrence dataset*

Data analysis was carried out for clean dataset of 359 data points. The statistical summary (Table 2) indicates that tamarind experiences a broad range of environmental conditions across its global distribution range.

3.3. *Principal component analysis global distribution data*

Investigation of the variable weighting scores of the first two components in the bioregional PCA analysis shows component 1 (24% of the variance) to be related to a pronounced dry period/water availability (i.e. moisture availability index seasonality, precipitation driest quarter and relative humidity warmest quarter). Component 2 (14% of the variance) is related to soil characteristics (i.e. organic carbon, total nitrogen, percentage sand, effective Cation Exchange Capacity) and mean temperature.

The PCA run for the biogeographical analysis shows a clustering of tamarind presence points within the first 2 components (Fig. 2). Component 1 distinguishes the Neotropical and Nearctic populations of tamarind from those of Afrotropical and Australian regions. Points found in the Indomalayan region appear to transect most of the range of the component 1, although there appear to be two

Table 2
Statistical summary for climatic variables.

Environmental Variable	Minimum	Maximum	Mean	Std. deviation	Environmental variable	Minimum	Maximum	Mean	Std. deviation
Mean annual temperature (°C)	15.80	28.60	24.66	2.17	Mean moisture availability index of coldest quarter	0.00	1.00	0.62	0.38
Mean diurnal range (°C)	5.80	19.60	11.24	2.17	Moisture availability index seasonality	0.00	122.00	47.05	30.28
Isothermality (°C)	4.20	8.90	6.48	0.93	Mean moisture availability index of warmest quarter	0.01	1.00	0.74	0.25
Temperature seasonality	26.90	471.70	181.44	87.18	Maximum monthly moisture availability index	0.10	1.00	0.99	0.08
Maximum temperature of warmest month (°C)	25.90	41.20	33.33	2.85	Mean moisture availability index of highest quarter	0.06	1.00	0.99	0.09
Minimum temperature of coldest month (°C)	4.20	24.10	15.78	3.35	Mean annual moisture availability index	0.03	1.00	0.74	0.19
Temperature range (°C)	7.50	30.50	17.55	4.00	Minimum monthly moisture availability index	0.00	1.00	0.27	0.35
Mean temperature wettest quarter (°C)	14.20	31.10	25.33	2.27	Mean moisture availability index lowest quarter	0.00	1.00	0.33	0.36
Mean temperature driest quarter (°C)	13.60	29.00	23.41	2.96	Bulk density subsoil (g/cm^{-3})	0.80	1.76	1.40	0.14
Mean temperature warmest quarter (°C)	18.00	32.60	26.83	2.41	Bulk density topsoil (g/cm^{-3})	0.76	1.65	1.36	0.14
Mean temperature coldest quarter (°C)	13.50	27.00	22.30	2.65	Percentage clay subsoil (%v/v)	3.75	68.00	31.31	13.57
Annual precipitation (mm)	32.00	3817.00	1333.25	635.70	Percentage clay topsoil (%v/v)	4.00	60.00	26.25	11.63
Precipitation of wettest month (mm)	9.00	1373.00	257.28	135.44	Percentage sand subsoil (%v/v)	3.00	89.00	39.06	15.96
Precipitation of driest month (mm)	0.00	190.00	18.49	30.66	Percentage sand topsoil (%v/v)	7.00	90.00	45.68	15.15
Precipitation seasonality (mm)	10.00	151.00	78.99	25.65	Percentage silt subsoil (%v/v)	4.00	53.00	23.71	8.47
Precipitation of wettest quarter (mm)	15.00	2902.00	650.43	315.79	Percentage silt topsoil (%v/v)	5.00	55.00	28.18	8.57
Precipitation of driest quarter (mm)	0.00	599.00	68.40	102.96	Percentage gravel topsoil (%v/v)	3.00	41.00	9.12	6.30
Precipitation of warmest quarter (mm)	3.00	1240.00	323.25	200.38	Percentage gravel subsoil (%v/v)	1.00	3.500	8.87	5.97
Precipitation of coldest quarter (mm)	0.00	1829.00	180.49	244.01	Effective CEC subsoil (cmolc kg^{-1})	2.00	59.00	17.83	10.46
Relative humidity of coldest quarter (%)	19.63	90.33	69.02	12.32	Effective CEC topsoil (cmolc kg^{-1})	3.00	55.00	16.94	9.84
Relative humidity of driest quarter (%)	17.70	89.40	62.99	14.27	Organic carbon content subsoil (% by weight)	0.20	1.91	0.50	0.27
Relative humidity warmest quarter (%)	30.83	90.70	66.12	11.75	Organic carbon content topsoil (% by weight)	0.35	7.00	1.20	0.82
Relative humidity wettest quarter (%)	36.23	90.70	76.71	6.22	pH subsoil	4.60	8.90	6.33	0.93
Mean sunshine duration coldest quarter (%)	18.33	91.37	62.56	14.21	pH topsoil	4.40	9.00	6.25	0.83
Mean sunshine duration driest quarter (%)	18.87	93.57	67.32	12.73	Total nitrogen subsoil (% by weight)	0.02	0.18	0.06	0.02
Mean sunshine duration warmest quarter (%)	21.80	85.43	59.63	9.96	Total nitrogen topsoil (% by weight)	0.03	0.34	0.11	0.06
Mean sunshine duration wettest quarter (%)	21.80	80.67	49.82	9.23					

Climate datasets were acquired from Worldclim (Hijmans et al., 2004, 2005) and the Climate Research Unit (New et al., 2000) datasets. Quarterly variables for relative humidity and sunshine were create based on data from New et al. (2000). Soil datasets with continuous values were produced by combining dominant soil type data and topsoil texture data from the TERRSTAT database (FAO, 2002a) with ISRIC WISE Derived Soil Parameter Data Set (ver2.2) (Batjes, 2002). A monthly water balance model was developed to produce Moisture availability index datasets, based on the method used in Global Agro-Ecological Zones project (FAO/IIASA, 2000). A Reference Evapotranspiration (New et al., 2000; IMWI, 2005), “Maximum Available Soil Moisture” (FAO, 2002b) and monthly rainfall datasets (Hijmans et al., 2004) were used as input data. Actual Evapotranspiration (Eta) was calculated according to the FAO Irrigation and Drainage paper 33. The Monthly Reference Evapotranspiration (Eto) dataset was reclassified based on figures from the “Soil Water Depletion Fraction (p) for Crop Groups and Maximum Evaoptranspiration (ETm)” (FAO, 1979). Tamarind was assumed to be within group three due to being most commonly found growing in the same agro-ecological conditions under non-irrigated conditions as other crops within this group (FAO, 1979).

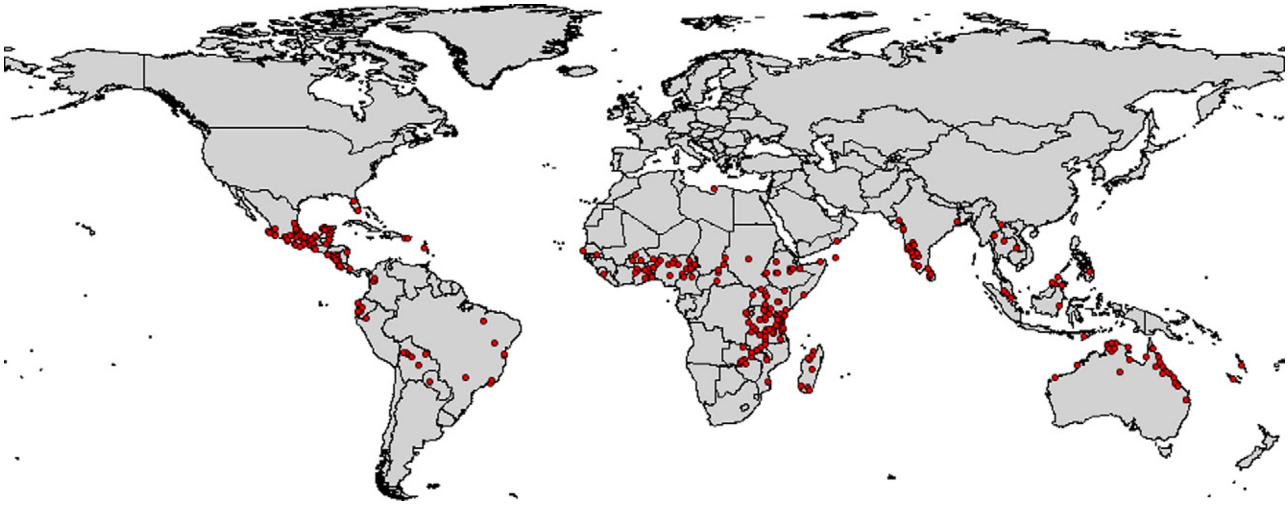

Fig. 1. Clean tamarind dataset (359 occurrence records).

main clusters within this biogeographical region. The points clustering with high values for component 1 (nearest the Neotropical specimens) are mainly from Southeast Asia, while those nearest the Afrotropical are mainly from South Asia. Based on the variable component weightings, the Afrotropical, Australian and Indomalayan – South Asia populations appear to experience drier conditions and a more pronounced dry period.

Component 2 also shows some separation of between the Afrotropical specimens and the Neotropical specimens. A number of Afrotropical specimens experiencing higher temperatures and more nutrient poor soils than those experienced by many of the Neotropical specimens. The Indomalayan population is constricted on component 2 and not found in areas of low temperature and very nutrient rich, high organic soil. The Indomalayan – Southeast Asia population is experiencing high temperature and high rainfall conditions in nutrient poor sandy soils. The Australian population appears to experience similar conditions to the Afrotropical population, while the Oceanic population is found in similar conditions to those of the main cluster from Indomalayan – Southeast Asia.

3.3.1. *Principal component sub-regional analysis*

The PCA run for the sub-regional analysis shows components 1 and 2 (explaining 23% and 16% of the variance respectively) to distinguish between the East and West Africa populations (Fig. 3). Component 1 is related to seasonality in rainfall and moisture availability. Component 2 is related to extreme temperatures and nutrient availability. As per the variable component weightings in West Africa the species is experiencing a highly pronounced dry period and high temperature conditions with sandy, nutrient/organic poor soil. In East Africa the species is found in both conditions of high and low moisture availability, though the latter only applies where the temperature is low and soil quality is good. The Southern Asia population appears to be spread across the full extent of component 1 however the Indian population is restricted

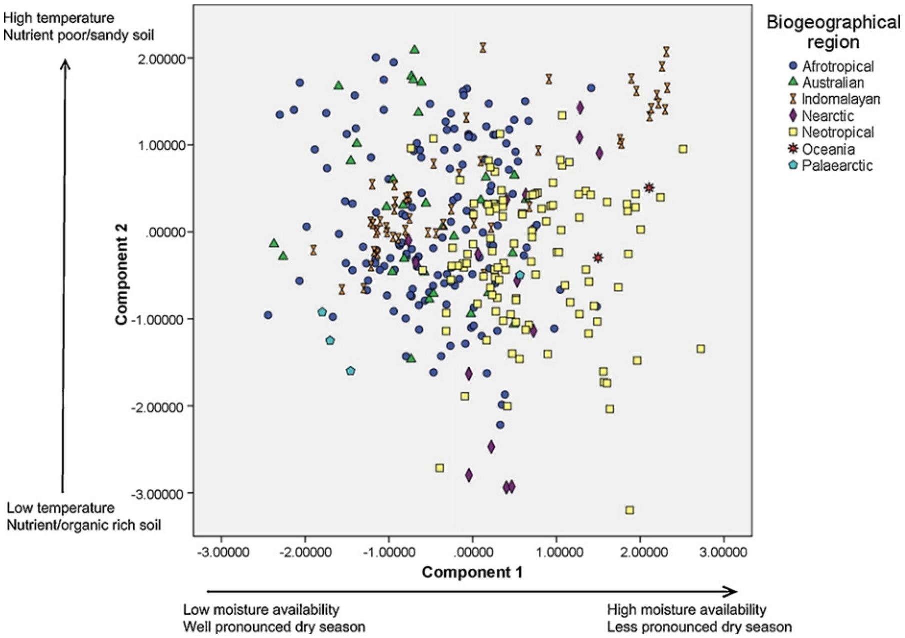

Fig. 2. Principle component analysis plot of component scores for components 1 and 2 classified by biogeographical region.

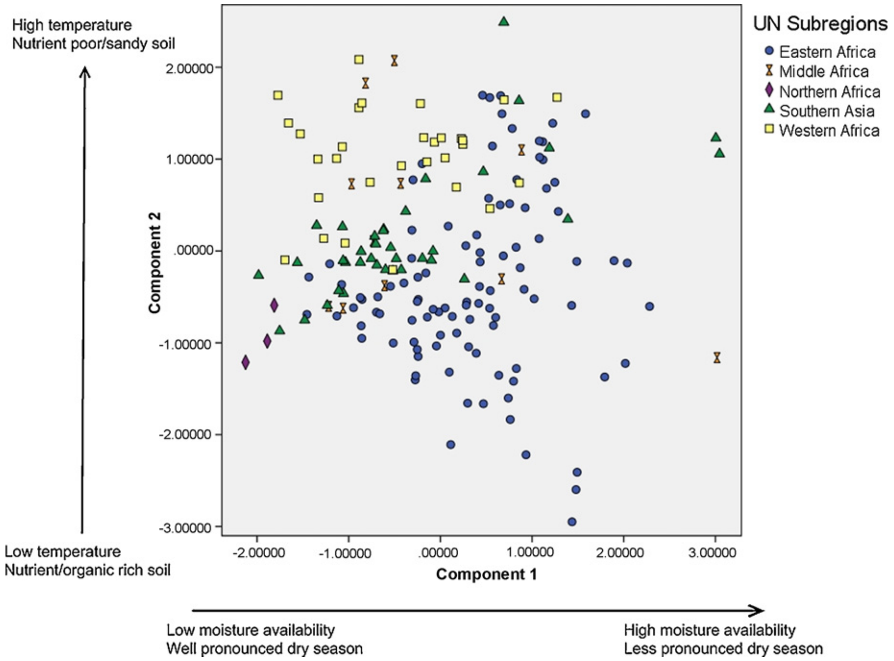

Fig. 3. Principle component analysis plot of component scores for components 1 and 2 classified by sub-region.

to areas of low moisture availability but only where temperatures are relatively low and soil relatively rich in organic material.

4. Discussion

The exploratory analysis of the clean dataset showed tamarind occurring in a broad range of both climatic and edaphic conditions. Both at the bioregional and sub-regional scale, tamarind has been found to experience different conditions in different regions. This indicates the existence of spatial non-stationarity across tamarinds global distribution. Spatial non-stationarity may be due to variation of environmental conditions available to the species, difference in community structure leading to changes in biotic interactions, phenotypic plasticity (Schluter, 2000; Peterson and Holt, 2003) or it could indicate the possibility of regional subpopulations or ecotypes within the main population brought about by differentiation of the niche (niche evolution) (Peterson and Holt, 2003).

Genetic variation is likely to exist for plant species, that cover large and/or environmentally heterogeneous distribution areas (Lillesø et al., 2001). Experimental work in tree species has revealed high among-population levels of genetic variation for quantitative traits related to adaptation and geographic structuring of the variation along climatic gradients (Howe et al., 2003; Savolainen et al., 2007; Aitken et al., 2008). As an increased number of range-wide diversity studies are conducted within tropical regions, many fruit tree species (including tamarind) with wide geographical ranges have shown a clear genetic structure with relatively high variation between populations both at the stand and country scale (Harris et al., 1997; Dawson and Powell, 1999; Diallo et al., 2007; Jamnadass et al., 2009). Cuni Sanchez et al. (2010) identified what appeared to be spatial genetic structuring for drought adaptive traits in the underutilised fruit tree species baobab (*Adansonia digitata*) in Benin. Human selection is also likely to lead to genetic differentiation of the niche. It had been suggested that artificial selection pressure during domestication (Miller and Knouft, 2006) and a high level of human introduction (Peterson and Holt, 2003) increases the likelihood of evolutionary effects in the form of geographic variation in niche characteristics. This leads to geographical subpopulations. With tamarind's broad distribution, semi-managed and domesticated nature and number of introductions, indicate that the likelihood of geographical variation within the niche and existence ecotypes is high.

The bioregional analysis showed differentiation in conditions experienced between the Afrotropical, Australian and Indomalyan-South Asia population and those of Neotropical, Oceanic and Southeast Asia regions. The Afrotropical, Australian and Indomalyan-South Asia populations were experiencing a more pronounced dry period. Tamarind is thought to be naturalised in Africa and the Indian subcontinent (NAS, 1979; El-Siddig et al., 2006). It is presumed that Indian Ocean traders carried the tree from India to Southeast Asia (Ochse and Bakhuizen Van Den Brink 1980 in Blench, 2008) and it was brought from Asia across the Pacific to the Americas with the Spanish (Patino, 1969). It would appear that as tamarind has been introduced into new areas it has expanded its range from the drier regions of Africa and South Asia to the wetter conditions of Southeast Asia and the Neotropics. Diallo et al. (2007) identified the Thailand populations to be most genetically similar to those in Guadeloupe (the only population sampled from Latin America) suggesting a source of introduction.

The Australian population was found in dry arid conditions, similar to those found in Africa. However tamarinds' introduction to Australia is often associated with the Macassarese people from South Sulawesi, Indonesia, some time in the last 1500 years (Worsley, 1955). If this is the case, tamarind appears to have carried out rapid adaptation to much drier conditions of Australia in comparison to its source of introduction, Indonesia. However, it may be the case that tamarind was pre-adapted to these drier conditions based on its native range in Africa (Meimberg et al., 2006; Ward et al., 2008).

The sub-regional scale analysis showed tamarind to be experiencing differing conditions within Africa. In West Africa the species is experiencing less rainfall and a more pronounced dry period with less nutrient/organic rich soil than found in East Africa. High genetic differentiation between East and West Africa has been found in many tree species (Harris et al., 1997; Dawson and Powell, 1999; Muchugi et al., 2006), including tamarind (Diallo et al., 2007). Lowe et al. (2000) identified variation at the country scale in bush mango (*Irvingia gabonensis* and *I. wombolu*), which he ascribed to limited geneflow derived from poor viability of the seeds and the historical effects of Pleistocene refugia. Pleistocene dynamics have influenced both the distribution and intraspecific portioning of diversity in temperate tree species, resulting in populations becoming isolated and differentiated in terms of their adaptive traits (Comes and Kadereit, 1998). Historical events may have a similarly strong influence on long lived tree species in the tropics (Newton et al., 1999; Lowe et al., 2000; Muchugi et al., 2006). However in the case of tamarinds whose seeds viable for months such an explanation is unlikely. Little is known about tamarind's association with natural vegetation making hypotheses about refugia difficult. Diallo et al. (2007) speculated that differentiation in tamarinds' neutral makers may be due to the Rift Valley preventing gene flow between both in the form of seed dispersion and those induced by human exchange and migration. We propose this may have also lead to a differentiation in the adaptive traits and therefore the niche of the species.

In this study, we suggest that as tamarind has expanded its range, adaptation of its niche may also have occurred. It is likely that when tamarind was introduced into a new continent – Asia, Australia and South America – the species likely to have suffered from genetic bottlenecks, unless introductions were multiple and from a relatively broad genetic base (Meimberg et al., 2006). This may have affected the adaptation potential of these isolated small populations due to reduced individual fitness and increased the probabilities of population extinction brought about by inbreeding depression. Semi-domesticated trees are particularly vulnerable to inbreeding depression in small isolated populations as trees often carry a heavy genetic load of deleterious recessive alleles (Atta-Krah et al., 2004; Aitken et al., 2008; Jamnadass et al., 2009; Dawson et al., 2009). Diallo et al. (2007) tried to identify the origin of tamarind origin, but was unable to do so due to a lack of clear bottlenecks in neutral markers between Africa and Asia. This may be due to the length of time tamarind has existed in India to where it is described as being "spread at an early date" (Burkill, 1997 in Blench, 2008). This period of time may have allowed for mutation to counter the founder effects. Several theoretical studies have shown that epistasitic and/or dominance variance can be converted to additive genetic variance as populations pass through bottlenecks. Thus, predicted reductions in additive genetic variance of bottlenecked invasive populations may be limited, or even reversed, allowing them to maintain their adaptive capacity during colonisation (Lande and Shannon, 1996; Naciri-Graven and Goudet, 2003; Müller-Schärer and Steinger, 2004; Reznick and Ghalambor, 2001).

The influence of population size and effect of inbreeding on adaptation potential of underutilised tree species could also have an influence on conservation; breeding and domestication strategies for underutilised tree species (see Dawson et al., 2009 and Atangana et al., 2010). Populations of these species on farms often suffer from the effects of isolation and genetic bottlenecks and are therefore vulnerable to inbreeding depression or genetic dilution from local stands. Jamnadass et al. (2009) and Dawson et al. (2009) have also highlighted that as trees are domesticated there is a paradox between decreasing molecular variation (measured by neutral markers) and increasing morphological diversity with increased human selection. Similar mechanisms may be acting on adaptive phenotypic characteristic which may have an influence on sustainability of the populations under changing climates.

In this study we have shown that the tamarind niche appears to vary across its distribution range, requiring issues of spatial non-stationarity to be considered during the modelling process. Tamarind's broad distribution, semi-domesticated nature and the genetic structure identified within the population mean that the likely effect of the regional variation may be due to evolutionary effects in the niche characteristics; resulting in geographical sub-populations. Peterson et al. (1999) and Peterson and Holt (2003) noted that use of ecological niche models to predict future distributions based on global climate change, depends fundamentally on whether niches are relatively stable over the geographic extent of a species range and over time. If a species' fundamental niche is consistent across its entire population (phylogenetic conservation in niche characteristics) (Huntley et al., 1989; Beerling et al., 1995; Prinzing et al., 2001; Peterson and Holt, 2003; Peterson, 2003; Thuiller et al., 2005; Estrada-Pena et al., 2006), it will be more accurate to model the global distribution, incorporating the full response curve of the species and predicting this into future climate. To predict species distribution for tamarind, regional modelling or multi-scale modelling (Osborne and Suarez-Seoane, 2002; Estrada-Pena et al., 2006) may provide more robust models and more accurate predictions.

Acknowledgements

The authors also acknowledge the technical advice from Dr. Patrick Osborne as well as comments and advice from Dr. Pete Shaw, Dr. Kate Schreckenberg and Ms. Louise Glew, School of Civil Engineering and the Environment, University of Southampton. The authors would also like to thank the anonymous reviewers for their useful comments which led to improvement of the manuscript.

References

Ahmed, M.F.U., Rahman, S.M.L., 2004. Profile and use of multi-species tree crops in the homesteads of Gazipur District, Central Bangladesh. Journal of Sustainable Agriculture 24 (1), 81–93.

Aitken, S.N., Yeaman, S., Holliday, J.A., Wang, T., Curtis-McLane, S., 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evolutionary Applications 1, 95–111.

Atangana, A.R., Beaulieu, J., Khasa, D.P., 2010. Wild genetic diversity preservation in a small-sized first generation breeding population of Allanblackia floribunda (Clusiaceae). Tree Genetics & Genomes 6, 127–136.

Atta-Krah, K., Kindt, R., Skilton, J.N., Amaral, W., 2004. Managing biological and genetic diversity in tropical agroforestry. Agroforestry Systems 61, 183–194.

Barnett, V., Lewis, T., 1978. Outliers in Statistical Data. Wiley, pp. 604.

Barry, S., Elith, J., 2006. When things go wrong: error and uncertainty in habitat models. Journal of Applied Ecology 43, 413–423.

Batjes, N.H., 2002. Soil parameter estimates for the soil types of the world for use in global and regional modelling (version 2.1). In: ISRIC Report 2002/02c.

Beerling, D.J., Huntley, B., Bailey, J.P., 1995. Climate and distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surface. Journal of Vegetation Science 6, 269–282.

Blench, R., 2008. A history of fruits on the Southeast Asian mainland. In: Occasional Paper 4 Linguistics, Archaeology and the Human Past. Indus Project Research Institute for Humanity and Nature, Kyoto, Japan.

Burrough, P.A., McDonnell, R.A., 1998. Principles of Geographical Information Systems. Oxford University Press, Oxford, UK.

Chanda, K., 2001. Evaluation of Tamarind (*Tamarindus indica* L.) Genotypes for productivity and grafting. Thesis. University of Agricultural Sciences, Dharwad.

Chapman, A.D., 1999. Quality control and validation of point-sourced environmental resource data. In: Lowell, K., Jaton, A. (Eds.), Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources, pp. 409–418.

Chapman, A.D., 2004. Internal Report No. 5 Environmental Data Quality, CIRA.

Codjia, J.T.C., Assogbadjo, A.E., Ekue, M.R.M., 2003. Diversity and local valorisation of vegetal edible products in Benin. Cahiers Agricultures 12 (5), 321–331.

Comes, H.P., Kadereit, J.W., 1998. The effect of quaternary climatic changes on plant distribution and evolution. Trends in Plant Science 3 (11).

Cuni Sanchez, A., Haq, N., Assogbadjo, A.E., 2010. Variation in baobab (Adansonia digitata L.) leaf morphology and its relation to drought tolerance. Genetic Resources and Crop Evolution 57, 17–25.

Dawson, I.K., Powell, W., 1999. Genetic variation in the Afromontane tree *Prunus africana*, an endangered medicinal species. Molecular Ecology 8, 151–156.

Dawson, I.K., Guarino, L., Jaenicke, H., 2007. Underutilised Plant Species: Impacts of Promotion on Biodiversity, International Centre for Underutilised Crops, Position Paper No. 2.

Dawson, I.K., Lengkeek, A., Weber, J.C., Jamnadass, R., 2009. Managing genetic variation in tropical trees: linking knowledge with action in agroforestry ecosystems for improved conservation and enhanced livelihoods. Biodiversity and Conservation 18, 969–986.

Diallo, B.O., Joly, I.H., Hossaert-McKey, M., McKey, D., Chevallier, M.H., 2007. Genetic diversity of *Tamarindus indica* populations: any clues on the origin from its current distribution? African Journal of Biotechnology 6 (7), 853–860.

Drummond, S.T., Sudduth, K.A., Joshi, A., 2000. Predictive ability of neural networks for site-specific yield estimation. In: The Second International Geospatial Information in Agriculture and Foresty Conference, Lake Buena Vista, FL, 10–12 January.

El-Siddig, K., Ebert, G., Lüdders, P., 1999. Tamarind (*Tamarindus indica*): a review on a multipurpose tree with promising future in the Sudan. Journal of Applied Botany 73, 202–205.

El-Siddig, K., Gunasena, H.M.P., Prasad, B.A., Pushpakumara, D.K.N.G., Ramana, K.V.R., Vijayanand, P., Williams, J.T., 2006. Tamarind, *Tamarindus indica L.* Southampton Centre for Underutilised Crops, Southampton, UK.

ESRI, 2000a. Arc View 3.2a.

ESRI, 2000b. Arc View – World data.

Estrada-Pena, A., Venzal, J.M., Sanchez Acedo, C., 2006. The tick *Ixodes ricinus*: distribution and climate preferences in the western Palaerctic. Medical and Veterinary Entomology 20, 189–197.

FAO, 1979. Yield Response to Water, vol. 33. Food and Agriculture Organisation of the United Nations, Rome.

FAO, 2002a. TERRSTAT Global Land Resources GIS Models and Databases and Land Degradation Assessment, Beta Version 1.0. Land and Water Digital Media Series 20. Food and Agriculture Organisation of the United Nations, Rome.

FAO, 2002b. Maximum Available Soil Moisture. FAO – Geo-Network.

FAO/IIASA, 2000. Global Agro-Ecological Zones (Global AEZ). Food and Agriculture Organisation of the United Nations, Rome.

GNS – National Geospatial-Intelligence Agency, 2005. U.S. National Imagery and Mapping Agency's (NIMA) Database of Foreign Geographic Feature Names – Country Files.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C., Peterson, T.A., 2004. New developments in museum based informatics and applications in biodiversity analysis. Trends in Ecology and Evolution 19, 497–503.

Hall, J.B., O'Brien, E.M., 2002. Ecology. In: Hall, J.B., O'Brien, E.M., Sinclair, F.L. (Eds.), Sclerocarya birrea, A Monograph. School of Agricultural and Forest Sciences, Publication Number 19, University of Wales Bangor, p. 157.

Harris, S.A., Fagg, C.W., Barnes, R.D., 1997. Isozyme variation in *Faidherbia albida* (Leguminosae, Mimosoideae). Plant Systematics and Evolution 207, 119–132.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2004. The WorldClim Interpolated Global Terrestrial Climate Surfaces. Version 1.3.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978.

Howe, G.T., Aitken, S.N., Neale, D.B., Jermstad, K.D., Wheeler, N.C., Chen, C.H.H., 2003. From genotype and phenotype: unraveling the complexities of cold adaptation in forest trees. Canadian Journal of Botany 81 (12), 1247–1266.

Huntley, B., Bartlein, P.J., Prentice, I.C., 1989. Climate control of the distribution and abundance of beech (*Fagus* L.) in Europe and North America. Journal of Biogeography 16, 551–560.

IMWI, 2005. Penman-Montieth reference evapo-transpiration dataset. In: World Water and Climate Atlas. International Water Management Institute.

Jamnadass, R., Lowe, A., Dawson, I., 2009. Molecular markers and the management of tropical trees: the case of indigenous fruits. Tropical Plant Biology 2, 1–12.

Lande, R., Shannon, S., 1996. The role of genetic variation in adaptation and population persistence in a changing environment. Evolution 50 (1), 434–437.

Lillesø, J.P.B., Dhakal, L.P., Shrestha, T.B., Nayaju, R.P., Shrestha, R., Kjaer, E.D., 2001. Tree planting zones in Nepal – an ecological approach based on vegetative types. In: DFSC Case Study No.1 TISC Technical Paper No. 103. Danida Forest Centre, Humlebæk.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G., Williams, P.H., 2003. Avoiding the pitfalls of using species distribution models in conservation planning. Conservation Biology 17, 1591–1600.

Louw, J.H., Scholes, M.C., 2006. Site index functions using site descriptors for *Pinus patula* plantations in South Africa. Forest Ecology and Management 225 (1-3), 94–103.

Lowe, A.J., Gilles, A.C.M., Wilson, J., Dawson, I.K., 2000. Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis. Molecular Ecology 9, 831–841.

McGarigal, K., Cushman, S., Stafford, S., 2000. Multivariate Statistics for Wildlife and Ecology Research. Springer–Verlag New York, Inc.

Meimberg, H., Hammond, J.I., Jorgensen, C.M., Park, T.W., Gerlach, J.D., Rice, K.J., McKay, J.K., 2006. Molecular evidence for an extreme genetic bottleneck during introduction of an invading grass to California. Biological Invasions 8, 1355–1366.

Miller, A.J., Knouft, H.J., 2006. GIS-based characterization of the geographic distributions of wild and cultivated populations of the Mesoamerican fruit tree, *Spondias purpurea* (Anacardiaceae). Journal of American Botany 93, 1757–1767.

Misra, M.K., Dash, S.S., 2000. Biomass and energetics of non-timber forest resources in a cluster of tribal villages on the Eastern Ghats of Orissa, India. Biomass and Bioenergy 18 (3), 229–247.

Muchugi, A., Lengkeek, A.G., Kadu, C.A.C., Muluvi, G.M., Njagi, E.N.M., Dawson, I.K., 2006. Genetic variation in the threatened medicinal tree *Prunus africana* in Cameroon and Kenya: implications for current management and evolutionary history. South African Journal of Botany 72 (4), 498–506.

Müller-Schärer, H., Steinger, T., 2004. Predicting evolutionary change in invasive, exotic plants and its consequences for plant–herbivore interactions. In: Ehler, L.E., Sforza, R., Mateille, T. (Eds.), Genetics, Evolution and Biological Control. CAB International.

Muok, B.O., Owuor, B., Dawson, I., Were, J., 2000. The potential of indigenous fruit trees: results of a survey in Kitui District, Kenya. Agroforestry Today 12 (1), 13–16.

Naciri-Graven, Y., Goudet, J., 2003. The additive genetic variance after bottlenecks is affected by the number of loci involved in epistatic interactions. Evolution 57 (4), 706–716.

NAS, 1979. Tropical legume Research for the Future, vols. 117–121. National Academy of Sciences, Washington, DC, pp. 112.

New, M., Lister, D., Hulme, M., Makin, I., 2000. A high resolution data set of surface climate over global land areas. Climate Research 21, 1–25.

Newton, A.C., Allnutt, T.R., Gillies, A.C.M., Lowe, A.J., Ennos, R.A., 1999. Molecular phylogeography, intraspecific variation and the conservation of tree species. Tree 14 (4), 140–145.

Osborne, P.E., Suarez-Seoane, S., 2002. Should data be portioned spatially before building large-scale distribution models. Ecological Modelling 157, 249–259.

Padulosi, S., Eyzaquirre, P.B., Hodgkin, T., 1999. challenges and strategies in promoting conservation and use of neglected and underutilised crop species. In: Janick, J. (Ed.), Perspectives on New Crops and New Uses. ASHA Press, Alexandria, VA.

Padulosi, S., Hodgkin, T., Williams, J.T., Haq, N., 2003. Underutilised Crops: Trends, Challenges and Opportunities in the 21st Century. CABI-IPGRI.

Park, S.J., Hwang, C.S., Velk, P.L.G., 2005. Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. Agricultural Systems 85, 59–81.

Patino, V.M., 1969. Plantes cultivadas y nnimals domesticos en America equinoccial. Tomo IV. Plamtes introducidas Imprenta Departamental, Cali, Colombia, pp. 233–235.

Peterson, T.A., 2003. Predicting the geography of species' invasions via ecological niche modeling. The Quarterly Review of Biology 78, 413–433.

Peterson, T.A., Holt, R.D., 2003. Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. Ecology Letters 6, 774–782.

Peterson, T.A., Soberon, J., Sanchez Acedo, C., 1999. Conservatism of the ecological niche in evolutionary time. Science 285, 1265–1267.

Polhill, D., 1988. Flora of Tropical East Africa: Index of Collecting Localities. Royal Botanical Gardens, Kew.

Pope, G.V., Pope, D.G., 1998. Flora Zambesiaca, Collecting Localities in the Flora Zambesiaca Area.

Prinzing, A., Durka, D., Klotz, S., Brandl, R., 2001. The niche of higher plants: evidence of phylogenetic conservation. Proceedings of the Royal Society of London Series B 268, 2383–2389.

Reznick, D.N., Ghalambor, C.K., 2001. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. Genetica 112–113, 183–198.

Savolainen, O., Pyhajarvi, T., Knurr, T., 2007. Gene flow and local adaptation in trees. Annual Review of Ecology and Systematics 38, 595–619.

Schluter, D., 2000. Ecological character displacement in adaptive radiation. American Naturalist 156, S4–S16.

Stape, J.L., Ryan, M., Binkley, D., 2004. Testing the utility of the 3-PG model for growth of *Eucalyptus grandis* × *urophylla* with natural and manipulated supplies of water and nutrients. Forest Ecology and Management 193, 219–234.

Stockwell, D., Peterson, T.A., 2001. Controlling bias in biodiversity data. In: Scott, M.J., Heglund, P.J., Morrison, M.L., Haufler, J.B., Rapheal, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrence: Issues of Accuracy and Scale. Island Press, Washington Covelo, London.

Sudjaroen, Y., Haubner, R., Wurtele, G., Hull, W.E., Erben, G., Spiegelhalder, B., Changbumrung, S., Bartsch, H., Owen, R.W., 2005. Isolation and structure elucidation of phenolic antioxidants from tamarind (*Tamarindus indica* L.) seeds and pericarp. Food and Chemical Toxicology 43, 1673–1682.

Thuiller, W., Richardson, M.D., Pysek, P., Midgley, G.F., Hughes, G.O., Rouget, M., 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology 11, 2234–2250.

Udvardy, M.D.F., 1975. A classification of the biogeographical provinces of the world. In: IUCN Occasional Paper No. 18. IUCN, Morges, Switzerland.

United Nations Statistics Division, 2008. Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings. United Nations, Visited 01/06/2008 http://unstats.un.org/unsd/methods/m49/m49regin.htm.

Van Zonneveld, M., Koskela, J., Vinceti, B., Jarvis, A., 2009. Impact of climate change on the distribution of tropical pines in Southeast Asia. Unasylva 60, 24–29, 231/232.

Ward, S.M., Gaskin, J.F., Wilson, L.M., 2008. Ecological genetics of plant invasion: what do we know? Invasive Plant Science and Management 1, 98–109.

WCMC, 2001. Udvardy's Ecoregions. World Conservation Monitoring Centre, FAO – Geonetwork.

Wieczorek, J., Qinghua, G., Hijmans, R.J., 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. International Journal of Geographical Information Science 18, 745–767.

Williams, J.T., Haq, N., 2002. Global Research on Underutilized Crops. An Assessment of Current Activities and Proposals for Enhanced Cooperation, pp. 46.

Williams, J.T., Haq, N., 2003. Underutilised crops for food security and poverty alleviation. In: Presentation at the International Workshop on Underutilized Plant Species, 6–8 May 2003, Leipzig, Germany.

Williams, P.H., Margules, C.R., Hilbert, D.W., 2002. Data requirements and data sources for biodiversity priority area and selection. Journal of Bioscience 27, 327–338.

Worsley, P.M., 1955. Early Asian Contacts with Australia Past and Present, pp. 1–11.

Zerger, A., 2004. Gridsampler. CSIRO Sustainable Ecosystems.