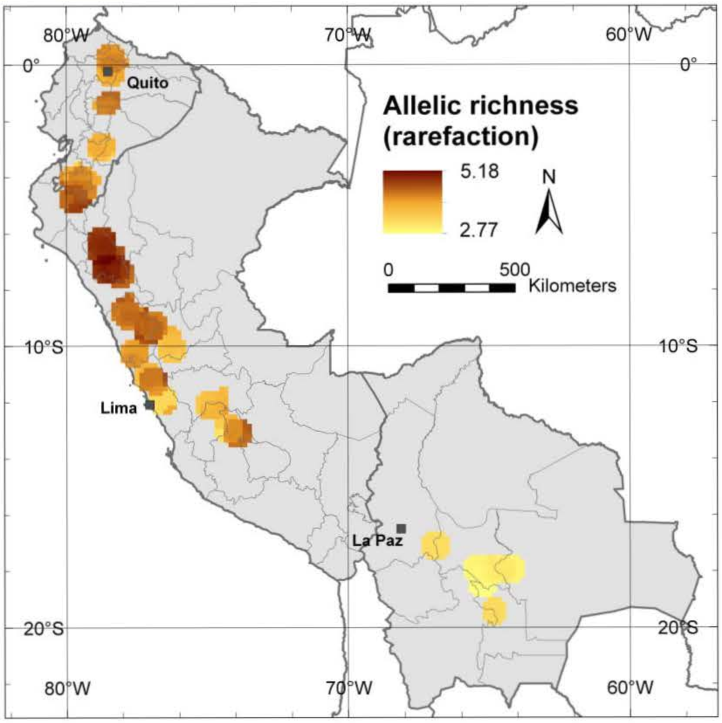
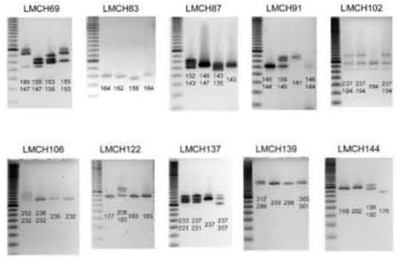
Economic plant diversity and distribution: application of spatial analysis for the conservation of germplasm diversity in Latin America

Maarten van Zonneveld





UNIVERSITEIT
GENT

FACULTEIT BIO-INGENIEURSWETENSCHAPPEN

Promotor: Prof. dr. ir. Patrick Van Damme
Faculty of Bioscience Engineering

Rector: Prof. dr. Paul Van Cauwenberge
Dean: Prof. dr. ir. Guido Van Huylenbroeck

Economic plant diversity and distribution: application of spatial analysis for the conservation of germplasm diversity in Latin America

by

Ir. Maarten van Zonneveld

Thesis submitted in fulfilment of the requirement for the degree of Doctor (PhD) in
Applied Biological Sciences

Cite as: Maarten van Zonneveld (2013) Economic plant diversity and distribution: application of spatial analysis for the conservation of germplasm diversity in Latin America. PhD thesis, Ghent University, Belgium.

Dutch translation of the dissertation title: Economische planten diversiteit and verspreiding: toepassing van ruimtelijke analyses ter bescherming van plantaardige genetische bronnen.

ISBN 978-90-5989-605-5

The author and the promotor give authorisation to consult, share and to copy parts of this work for personal use only. Any other use is limited by Laws of Copyright. Permission to reproduce any material contained in this work should be obtained from the author.

Prof. dr. ir. P. Van Damme	M. van Zonneveld
Promotor	Author

Faculty of Agricultural and Applied Biological Sciences
Department Plant Production
Laboratory of Tropical and Subtropical Agronomy and Ethnobotany
Coupure links 653
B-9000 Ghent
Belgium

Credits cover images: Potato species image: Ximena Cadima (PROINPA, Bolivia) – Peach palm image and allelic richness map: Maarten van Zonneveld (Bioversity International and Ghent University) – Cherimoya image: Xavier Scheldeman (Bioversity International) - DNA gel electrophoresis images: Iñaki Hormaza (IHSM la Mayora, Consejo Superior de Investigaciones Científicas, Spain).

Abbreviations and acronyms

AAIA	Allelic Aggregation Index Analysis
AFLP	Amplified Fragment Length Polymorphism
ANOVA	Analysis Of Variance
API	Application Programming Interface
APG	Angiosperm Phylogeny Group
ARS	Agricultural Research Service
AUC	Area Under Curve
BRAHMS	Botanical Research and Herbarium Management System
°C	Celsius degrees
CA	Correspondence Analysis
CAP	Canonical Analysis of Principal Coordinates
CATS	Conservation assessment tools
CART	Classification Categorical Response Trees
CBD	Convention on Biological Diversity
CCA	Canonical Correspondence Analysis
CCAFS	Climate Change Agriculture and Food Security
CEDERENA	Corporación para el Desarrollo de los Recursos Naturales
CGIAR	Consultative Group on International Agricultural Research
CGN	Centrum voor Genetische Bronnen, Nederland
CIAT	Centro Internacional de Agricultura Tropical
CIP	Centro Internacional de la Papa
CSIC	Consejo Superior de Investigaciones Científicas
CSS	Cascading Style Sheets
CWR	Crop Wild Relatives
DArT	Diversity Arrays Technology
DCA	Detrended Correspondence Analysis
df	degrees of freedom
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EEM	Environmental Envelope Modelling
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
ENFA	Ecological Niche Factor Analysis
EST	Expressed Sequence Tagged
F	fixation index
FAO	Food and Agriculture Organization of the United Nations
FGR	Forest Genetic Resources

FIGS	Focused Identification of Germplasm Strategy
g	gram
GADM	Global Administrative Areas
GARP	Genetic Algorithm for Rule-set Prediction
GBIF	Global Biodiversity Information Facility
GAM	General Additive Model
GCDT	Global Crop Diversity Trust
GCM	Global Circulation Model
GEF	Global Environmental Facility
GD	genetic distance
GIS	Geographic Information Systems
GLCF	Global Land Cover Facility
GLM	General Linear Model
GPA	Global Plan of Action
GPS	Global Positioning System
GRASS	Geographic Resources Analysis Support System
GRIN	Germplasm Resources Information Network
ha	hectare
He	expected heterozygosity per locus
HTML	Hyper Text Markup Language
ICRAF	World Agroforestry Centre
IHSM	Instituto de Hortofruticultura Subtropical y Mediterránea
ILWIS	Integrated Land and Water Information System
INCO	International co-operation activities
INIA (Peru)	Instituto Nacional de Innovación Agricola
INIA (Spain)	Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria
INFOR	Instituto Forestal
INTA	Instituto Nacional de Tecnología Agropecuaria
IPK	Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung
IPNI	International Plant Names Index
ITPGRFA	International Treaty on Plant Genetic Resources for Food and Agriculture
IUCN	International Union for Conservation of Nature
JSTOR	Journal Storage
km	kilometer
KML	Keyhole Markup Language
K	number groups in Bayesian clustering approaches
kv	kilovolt
LAFORGEN	Latin American Forest Genetic Resources Network
Maxent	Maximum entropy
masl	meters above sea level

MCMC	Markov Chain Monte Carlo
MgCl2	Magnesium chloride
min	minute
MLM	Modified Location Model
m	metre
mm	millimeter
MODIS	Moderate Resolution Imaging Spectroradiometer
n	number of individuals
Na	average number of alleles per locus
NASA	National Aeronautics and Space Administration
ng	nanogram
(NH4)2SO4	Ammonium sulphate
NGRP	National Genetic Resources Program
NMDS	Non-Metric Multidimensional Scaling
NOAA	National Oceanic and Atmospheric Administration
NDVI	Normalized Difference Vegetation Index
NPGS	National Plant Germplasm System
NUFFIC	Netherlands organisation for international cooperation in higher education
OECD	Organisation for Economic Co-operation and Development
p	probability
PAS	private alleles function
PCA	Principal Components Analysis
PCoA	Principal Coordinates Analysis
PGR	Plant Genetic Resources
PSTV	Potato Spindle Tube̊r Viroid
PVX	Potato Virus X
PVY	Potato Virus Y
QGIS	Quantum GIS
QTL	Quantitative Trait Locus
RDA	Redundancy Analysis
ROC	Receiver Operation Curve
RNA	Ribonucleic acid
s	seconds
SAGA	System for Automated Geoscientific Analyses
SDTF	Seasonally Dry Tropical Forests
SINGER	System-wide Information Network for Genetic Resources
SNP	Single Nucleotide Polymorphism
spp.	subset of species from a genus
SOTERLAC	Soil and Terrain database for Latin America and the Caribbean
SRTM	Shuttle Radar Topography Mission

SSR	Simple Sequence Repeat
subsp.	subspecies
UNEP	United Nations Environment Programme
UMA	Universidad de Málaga
US	United States
USDA	United States Department of Agriculture
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
var.	variety
VMABCC	Viceministerio de Medio Ambiente, Biodiversidad y Cambios Climáticos
WGS	World Geodetic System
y	year
µl	microliter
µm	micrometer

Acknowledgements

There are several people that I would like to thank. First of all, dr. Xavier Scheldeman who encouraged me to start this Phd and has been a mentor for many years. I would like to acknowledge dr. Milena Holmgren who inspired me as MSc student to continue in biodiversity research. I am gratefull to my promotor prof. dr. ir. Patrick Van Damme for his supervision during the development of my dissertation.

I would like to thank Ximena Cadima, dr. Judy Loo, dr. ir. Evert Thomas, dr. Gea Galluzzi, dr. Ian Dawson, dr. José Hormaza, dr. ir. Jacob van Etten, Nora Castañeda, Fernando Patiño, Marcela Beltran, dr. Pilar Escribano, dr. María Viruel, Willman Garcia, César Tapia, José Romero and Manuel Sigueñas for their contributions to one or more chapters in this dissertation.

Special thanks to Vincent Johnson, dr. Marleni Ramirez, dr. ir. Ronald van den Berg, dr. Laura Snook and dr. ir. Marc Sosef whose comments and edits improved considerably specific parts of this thesis.

I would like to thank prof. dr. ir. Wim Verbeke, prof. dr. ir. Paul Goetghebeur, and prof. dr. ir. Rob de Wulf, and again dr. Milena Holmgren and dr. ir. Xavier Scheldeman for having accepted to be member of the board of examiners.

I am also grateful to Bioversity International and to the encouragements of my colleagues during the process of developing the thesis. Through my work in Bioversity, I was able to carry out this work.

Several people have helped in making specific case studies a success. Staff of the PROINPA Foundation and dr. Alberto Salas from the International Potato Center (CIP) for support in field work in Bolivia helped in realizing the study on potato wild relatives of which the results have been presented in chapter 4. Dr. Jorge Rojas and his team from PROINPA, Bernardo Guzmán from PROINPA, personal from INIA-Peru and Doris Chalampuente, Fernando Paredes, Marcelo Tacán, Eddie Zambrano and Edwin Naranjo from INIAP Ecuador helped in the sampling and logistics of the cherimoya genetic diversity study of which the results were presented in chapter 6.

Finally, I would like to thank my spouse Evelyn Quintana and my daughters Karla and Sophia for their patience and understanding during the preparation of this thesis. You have been a real inspiration to complete this work.

Summary

Plant genetic resources (PGR), i.e. inter- and intra-specific plant diversity for current and future human use- are vital for the survival and well-being of humanity. However, the economic plant diversity and its conservation status in natural stands, managed ecosystems, farmers' fields, home gardens and in other *in situ* settings is often poorly understood and regarded as seriously threatened by human disturbance. There is therefore an increasing recognition of the necessity to assess and optimize conservation actions and link these effectively with *ex situ* preservation approaches. Geographic Information Systems (GIS) and ecogeographic analysis could contribute significantly to improved understanding and monitoring of spatial and temporal patterns of biodiversity to support conservation actions of plant species. The hypothesis tested in this dissertation is the following: geospatial analysis of plant diversity and distribution can clearly detect geographic inter-specific and intra-specific diversity and distribution patterns, which allows prioritizing those plant species populations or distribution areas that should be considered for *in situ* conservation and germplasm collecting. This proposition is tested by (1) mapping plant diversity and distribution in several plant genetic resources case studies; and (2) by soliciting experts' feedback on the modelling of species' distributions. Their responses allow to better understanding of how useful such analyses can be in supporting local practitioners in the implementation of conservation measures.

The first part of this work presents an overview of those relevant techniques and advances in ecogeographic studies of PGR that can be used to analyse biodiversity data on the basis of field-collected data and to target further germplasm collecting for *ex situ* conservation. First of all, some important general considerations are articulated for setting up new research projects that are aimed at assessing the conservation status of PGR and/or monitoring trends in (economic) plant diversity on the basis of ecogeographic data. A brief introduction to commonly used methods and techniques for the analysis of inter- and intra-specific diversity is provided. The latter include multivariate methods such as clustering and ordination. Several techniques to map (economic) plant diversity data are discussed and ways to check and improve data quality are explained. Finally a synopsis of methods for Environmental Envelope Modelling (EEM- see below) and an overview of useful open-access and commercial statistical and GIS packages is presented.

Special emphasis is given to molecular marker concepts and examples of their application as well as geospatial analysis to carry out diversity analysis and optimize *in situ* conservation. Recent development of new powerful molecular tools that reveal many genome-wide polymorphisms has created novel opportunities for assessing genetic diversity, especially when these markers can be linked to key adaptive traits and are

employed in combination with new geospatial methods of geographic and environmental analysis. New methods to prioritize varieties, populations and geographic areas for *in situ* conservation, and to enable monitoring of genetic diversity over time and space, are now available to support *in situ* germplasm management of crop and tree genetic resources.

For most plant species in the tropics and subtropics, including many crop wild relatives and socio-economically important tree species, only a limited amount of information on their natural distribution is currently available. EEM is considered a useful tool for providing vital missing information on the natural distribution of a species. Nevertheless, application of EEM for conservation planning requires careful validation. Opinions of experts who have worked in the field on conservation, seed collection and ecology of the specific species of interest offer a valuable and independent information source to validate EEM, because of their first-hand experience with species occurrence and absence. However, their use in model validation has always remained limited due to the subjectivity of their feedback. In this thesis, cultural consensus theory is utilized to formalize expert model evaluations. Such approaches allow a wider use of this information in model validation and improvement, and complement conventional validation methods of presence-only modelling. Online GIS and survey applications facilitate expert consultation.

The concepts and methods described above are applied in three case studies to demonstrate their usefulness for PGR *in situ* conservation and germplasm collecting: (1) an assessment of the diversity and conservation status of potato wild relatives endemic to Bolivia; (2) a review of peach palm (*Bactris gasipaes* Kunth) diversity, distribution, and PGR management; and (3) mapping molecular intra-specific diversity of cherimoya (*Annona cherimola* Mill.).

Bolivia is a centre of wild relative diversity for several crops, among them potato, which is a globally significant staple and the principle food crop in this country. Despite their relevance for plant breeding, limited knowledge exists about their *in situ* conservation status. GIS and distribution modelling with the software Maxent are applied to better understand geographic patterns of endemic wild potato diversity in Bolivia. In combination with threat layers, the conservation status of all endemic species, 21 in total, is assessed. Following the International Union for Conservation of Nature (IUCN) Red List ecogeographic criteria area of occupancy and extent of occurrence, at least 71 % (15 of 21 species) has a preliminary vulnerable status or worse. Our results show that five of these species require special conservation attention because they are highly threatened by increased accessibility of areas for human disturbances, fires and livestock activities pressure leading to overgrazing. Highest species richness occurs in south-central Bolivia, in the departments Santa Cruz and Chuquisaca. However, this area is severely threatened by the menaces mentioned above. The costs to implement conservation measures at these locations may be too high compared to other areas. Therefore a prioritization exercise, excluding 25 % of the most-threatened occurrence sites, is carried out identifying the most species rich area and complementary areas. The first priority area for *in situ* conservation according to our reserve selection exercise is central Bolivia, Cochabamba, which is less threatened than the potato wild relatives' hotspot in south-central Bolivia. Only seven of the 21 species have been observed in protected areas. Understanding of the coverage of potato wild relatives' distribution in protected areas can be improved by starting inventories in parks and reserves with high levels of modelled diversity. Finally, five of the 21 are either not conserved in one or more genebanks across the world or are conserved with less than five accessions, i.e. samples of living plant material collected from particular locations. New materials of these species should be included in genebanks to improve *ex situ* conservation of the potato gene pool.

Peach palm (*Bactris gasipaes* Kunth) is a multi-purpose palm tree that produces starchy, edible fruits and palm hearts. It may be considered the most economically important domesticated palm species of the Neotropics and has been widely used since early pre-Columbian times. Wild and cultivated peach palm populations are genetically diverse and could offer useful traits for breeding. Changes in land use and climate change pose a serious threat to wild populations *in situ*. While several large *ex situ* field collections of cultivated peach palm accessions exist, these are increasingly difficult to maintain because of the high costs. Screening peach palm diversity for biochemical and morphological traits of commercial and nutritional value would provide a basis for rationalizing collections and enhance future use of peach palm genetic resources. Indeed, well-chosen elite material could then be used either directly for production, or in breeding to develop improved peach palm varieties. At the same time, better propagation techniques should be developed to ensure wide distribution of elite peach palm clonal material.

A case study with cherimoya (*Annona cherimola* Mill) in its Andean distribution range, explores the possibilities of incorporating molecular marker characterization data into Geographic Information Systems (GIS) to allow visualization and better understanding of spatial patterns of genetic diversity as a key input to optimize conservation and use of plant genetic resources, Cherimoya is a Neotropical fruit tree species. Its fruits are widely praised for their excellent taste and smell. The species is therefore considered to have high potential for commercial production and income generation for both small and large-scale producers in subtropical climates. Spatial analyses are utilized to (1) improve the understanding of spatial distribution of genetic diversity of cherimoya natural stands and cultivated trees in Ecuador, Bolivia and Peru characterized with microsatellite molecular markers (SSRs); and (2) formulate optimal conservation strategies by revealing priority areas for in situ conservation, and identifying existing diversity gaps in ex situ collections. High levels of allelic richness and locally common alleles are evidenced in cherimoya's putative centre of origin, southern Ecuador and northern Peru. This suggests accumulated genetic resources resulting from a long history of human management and adaptation of trees to local climate conditions. Therefor these areas should be prioritized for *in situ* conservation. Levels of diversity in southern Peru and especially in Bolivia are significantly lower. However, the tree stands in these areas belong to a genetically different population than those in southern Ecuador and northern Peru. They may contain genetic resources that are not present in the above-mentioned centres of diversity. It is therefore important to consider these areas too in defining conservation strategies.

The results obtained in the different case studies support the hypothesis of this dissertation as stated above, i.e. that geospatial analysis of plant diversity and distribution analysis can clearly detect geographic inter-specific and intra-specific diversity patterns, which allows more effective prioritization of those plant species populations and distribution areas to be considered for *in situ* conservation and germplasm collecting.

First, the results from the expert validation exercise of EEM show that many professionals are fairly positive about the use of distribution modelling for *in situ* conservation planning. This suggests that such analysis can support local professionals in their planning work of managing and conserving plant genetic resources. However, these tools have limitations in their precision. Consequently their applicability to support conservation actions in the field remains restricted. Further improvement of distribution modelling techniques to provide support for more local conditions would therefore be helpful. A key point is to increase the availability of more detailed geospatial environmental layers.

Many professionals working with plant genetic resources and/or in the field of economic botany may opt to apply themselves geospatial analysis but do not necessarily have the required experience. In parallel with the research for this thesis, a manual on plant diversity and distribution has been developed. This manual responds to the increasing demand from professionals working with plant genetic resources such as botanists, agronomists and ecologists for this type of analyses.

Secondly, in the case studies of mapping wild potatoes and cherimoya diversity clear and detailed geographic patterns of respectively taxonomic and molecular diversity could be detected. This suggests that diversity mapping allows more effective prioritizing of areas for conservation and germplasm collecting at both inter- and intra-specific levels. On the contrary, the peach palm review of existing genetic studies did not provide evidence for areas of high diversity of this particular species. Studies from existing literature included often only a limited number of populations and used different sampling methods and marker types. This makes it difficult to carry out meta-analysis. Therefore standardization of methods and range-wide analysis across species distributions are recommended for better detection of genetic diversity hotspots of plant species. Range-wide analyses of species genetic diversity become increasingly feasible as the cost of diversity studies with molecular markers becomes cheaper.

Such analyses can also be applied to other plant species. This has been done now for example in a distribution and conservation assessment of 100 tree species native to Latin America and the Caribbean named MAPFORGEN (www.mapforgen.org).

Several models and scenarios are discussed that can explain the diversity patterns that are observed in the wild potato and cherimoya studies. The drivers that shape the geographic patterns of diversity in these two case studies are completely different. The occurrence and evolution of potato wild relatives have been shaped by natural processes over hundreds, thousands and millions of years. The distribution of cherimoya genetic diversity follows a pattern of human-mediated crop dispersal that started after about 13,000 years ago when agriculture originated in different parts of the world during the Neolithic revolution.

Future challenges are being discussed including approaches and concepts of data sharing and standardization. The latter would make it possible to combine plant diversity datasets leading to stronger analyses to detect geographic patterns of plant diversity and distribution. Standardization of passport data, characterization and evaluation also enhance comparability of study results in meta-analyses. There is also an increasing need to learn more about phenotypic variation in adaptive characteristics and other functional traits of plant species to identify materials with traits of interest; to understand the responses of plant species to climate change; and to estimate the evolutionary potential of populations, to name a few. However phenotypic evaluation is costly. Smart approaches such as pre-selection of plant materials and populations by means of molecular characterization are required to optimize evaluation of functional genetic variation. The potential and limitations of using molecular characterization and citizen science in monitoring economic plant diversity are discussed. The thesis ends with reflections on possible environmental and cultural factors that influence maintenance, reduction or increase in cultivated plant diversity. A better understanding of these drivers helps to decide which interventions are necessary to enhance use and conservation of PGR under specific cultural, socio-economic, and biophysical conditions.

Contents

Introduction

Plant Genetic Resources (PGR), inter- and intra-specific plant diversity for current and future human use, are vital for the survival and well-being of humanity. Domesticated plant species are critical to global food security. Some crops are also of great importance for other purposes such as fibre or fodder production. In addition to these domesticates, many wild plants still play an important role in meeting local needs for food, fuel, medicine and construction materials. Crop Wild Relatives (CWR) are of interest for crop breeding programmes because they can contain traits of interest such as pest and disease resistance or tolerance to abiotic stresses (Hajjar and Hodgkin 2007). There are currently thousands of underutilized plant species and varieties displaying traits of interest to meet present and future needs. The value of many other plant species for human use is yet to be discovered and confirmed.

It is generally accepted that modernization of agriculture and land use changes negatively affect economic plant diversity maintained on farmers' fields and in home gardens (van de Wouw et al. 2010a). This might eventually lead to genetic erosion (van de Wouw et al. 2010a). As a result, many collecting missions have been organized in the past decades to establish extensive genebank collections for important food crops (Thormann et al. 2012). The genetic integrity of accessions, i.e. samples of living plant material collected from particular locations, is maintained as much as possible in *ex situ* genebank collections to conserve the specific characteristics of each material for evaluation, breeding and direct use. However, these *ex situ* collections do not maintain the continued process of interactions between plants, humans and environmental factors that take place in *in situ* settings (Altieri and Merrick 1987). This process is thought to be important to assure evolution of plant species with their environment and under human selection. Human-plant interactions are especially relevant for domesticates.

There is therefore the need to assess the diversity status and dynamics of PGR in *in situ* settings to prioritize and optimize *in situ* conservation actions and link these effectively with *ex situ* preservation approaches (Frankel et al. 1995a; Palmberge-Lerche 2008; Dulloo et al. 2010; FAO 2010a; 2011; Pereira et al. 2013). At the same time, these type of analyses are useful to identify remaining geographic gaps of diversity that are missing in existing genebank collections and that should therefore be targeted for germplasm collecting (FAO 2010a; 2011).

Following Article 2 of the Convention of Biological Diversity (CBD), an *in situ* setting means 'conditions where genetic resources exist within ecosystems and natural habitats, and, in the case of domesticated or cultivated species, in the surroundings where they have developed their distinctive properties.' (http://www.cbd.int/convention/articles/?a=cbd-02). Centres of plant diversity, domestication and/or diversification meet these conditions for respectively wild and domesticated species. These areas should thus be identified and prioritized for *in situ* conservation of economic plant diversity. In areas where agriculture started early in history, for example, cultivated plant species have been subjected to a long history of natural and human selection. This results in high (accumulated) levels of genetic resources. With respect to natural plant populations, some areas may have acted as refugia over long periods of times such as within glacial periods. In such areas, plant species could have maintained high levels of genetic variation in sufficiently large populations. From these areas, species could then have re-colonized areas and established themselves in new ones.

In situ settings occur in different domestication stages from natural vegetation to intensive cultivation including natural stands, managed ecosystems, home gardens and farmers' fields (Frankel et al. 1995a; Wiersum 1997; Clement 1999). Plant-man interactions differ across this continuum. Consequently, so is the level and type of plant diversity in these different settings. But these settings have in common that they are all dynamic environments (Frankel et al. 1995a). In all these habitats, the plant populations, their diversity and on-going selection processes can be maintained through *in situ* conservation (Frankel et al. 1995a).

In situ conservation is also the method of choice for species with recalcitrant seeds that cannot be stored for long periods in *ex situ* seed banks, and for plants whose biology (e.g. long time lapse to maturity, seed dormancy) makes human-managed regeneration costly or difficult. These latter features apply to thousands of tree species. In the case of lots of tropical and subtropical economic tree species that provide timber, fruit and other non-timber products, genetic resources are often principally or exclusively maintained in home gardens, on-farm and/or in natural populations. On-farm or in home garden conservation of tree species within their natural distribution ranges is often also referred to as *circa situm* conservation to distinguish this type of management from *in situ* conservation of tree species in natural populations (Boshier et al. 2004; Hollingsworth et al. 2005). The diversity of tree species maintained *circa situm* and *in situ* can be considered a treasure trove for yet uncharacterized and/or unknown traits for growers, breeders for different consumer markets and possibly also local people (Scheldeman et al. 2003; Ræbild et al. 2011). However, trees in modified natural habitats and farmland may be susceptible to particular pressures such as inbreeding depression as a consequence to decreased population sizes, limited inter-tree connectivity and global climate change (Dawson et al. 2009, 2011; Vranckx et al 2011). Particularly insect-mediated outcrossing woody perennials are sensitive to such pressures (Vranckx et al. 2011).

The main purpose of PGR *in situ* conservation is to maintain genetic variation in cultivated and natural plant species populations for phenotypic selection by farmers and/or natural processes (Cleveland and Soleri 2007). This allows maintenance of processes of (micro) evolution and continuous adaptation of crops or wild plants to their environments. The genetic structure of populations can change when phenotypic traits are heritable and selection is sufficiently strong. Following Darwin's concepts of selection, this allows cumulative directional genetic response over generations, i.e. micro-evolution of these populations to natural and human selection (Cleveland and Soleri 2007; Thomas et al. 2011). Micro-evolution in plant populations is further driven by factors such as random mutation, recombination and genetic drift (Briggs and Walters 1999). The weight of all these different factors to shape genetic variation within and between populations is still open to debate (Briggs and Walters 1999).

As an additional factor in on-farm conservation, many smallholders in all parts of the world introduce periodically new materials from neighbours and other localities into their systems to sustain productivity (Zeven 1999). These factors and activities together make on-farm PGR management a dynamic system of use of crop genetic diversity. Farmers may select for changing preferences as well choose to maintain desired phenotypic traits (Cleveland and Soleri 2007). The variety of traits that is maintained and evolving under farmers' care is often unknown to conventional breeders, entrepreneurs and consumers. This makes on-farm conservation areas potential sources of untapped diversity for the development of new crop varieties for local and wider use. Even genetic diversity itself in cultivated populations may be a trait of farmers' selection for ecosystem services such as pest and disease control (Hajjar et al. 2008).

The CBD, established in 1992, calls for a Global Strategy for Plant Conservation (CBD 2012). In addition to the CBD, the Global Plan of Action for the Conservation and Sustainable Utilization of Plant Genetic Resources for Food and Agriculture (GPA) was adopted in 1996 and updated in 2011 (FAO 2011). The International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) (FAO 2009), entered into force in 2004. Both the latter were initiated upon recognizing the potential of PGR and their importance for food security and sustainable agricultural production. The importance of PGR is also recognized in Target 13 of the 20 Aichi Biodiversity Targets that have been established in the Strategic Plan for Biodiversity 2011-2020 (CBD 2010). This specific target states that 'by 2020, the genetic diversity of cultivated plants and farmed and domesticated animals and of wild relatives, including other socio-economically as well as culturally valuable species, should be maintained, and strategies be developed and implemented for minimizing genetic erosion and safeguarding their genetic diversity' (CBD 2010).

These international frameworks act parallel to each other and aim to enhance the conservation and use of plant diversity. All these plans and agreements include a component that stresses the importance to increase information on biodiversity status and trends and formulate actions to enhance its conservation and use. This thesis contributes to this component and the results are relevant for Article 7 of the CBD that calls for the identification and monitoring of biodiversity, paying particular attention to those species and varieties that offer the greatest potential for sustainable use and require urgent conservation measures (CBD 1992). This dissertation also contributes to Aichi Biodiversity Target 19. This target states that ‘by 2020, knowledge, the science base and technologies relating to biodiversity, its values, functioning, status and trends, and the consequences of its loss, are improved, widely shared and transferred, and applied’ (CBD 2010). Also, the GPA formulates specific activities to increase information on PGR status and trends and defines actions to enhance its conservation and use. Priority Action 1 calls for increased surveying and inventorying of PGR for food and agriculture. Further, Priority Action 7 recommends planned and targeted collecting efforts of PGR for food and agriculture. The importance of these activities is further confirmed in Article 5 of the ITPGRFA. The latter article promotes activities related to conservation, exploration, collection, characterization, evaluation and documentation of PGR for food and agriculture. In addition to defining these priorities, each of these international frameworks emphasizes the need to strengthen local capacities to carry out research related to diversity and genetic resources conservation and sustainable use.

Geographic Information Systems (GIS) and ecogeographic analysis can contribute significantly to improve understanding and monitoring of spatial and temporal patterns of plant diversity (Guarino et al. 2002). Results obtained from spatial analyses allow the formulation and implementation of better-targeted, and hence more effective conservation strategies of inter and intra-specific plant diversity. Such studies are especially useful to evaluate current conservation status of plant species; prioritize areas for conservation; and monitor status and trends of existing plant diversity levels and the use of PGR by humans. GIS have also proven useful for establishing effective genebank management as they can be used to link climate and other ecogeographic data to georeferenced passport data (Guarino et al. 2002) This information helps to define core collections for more detailed morphological characterization and agronomic evaluation that include materials from different agro-ecological zones (Parra-Quijano et al. 2011). Preferable, this is combined with available molecular, morphological and biochemical characterization and/or phenotypic and agronomic evaluation data. GIS are also being used to identify geographic and environmental collection gaps (Maxted et al. 2008). GIS tools allow to carry out complex analyses combining different (spatial) data sources (Guarino et al. 2002). At the same time, GIS can be used to generate clear maps, which facilitate the uptake of outcomes of diversity status and trends assessments by the respective responsible authorities, and encourage development and implementation of conservation policies (Jarvis et al. 2010). In recent years, technological advances and increasing availability of powerful computers and GPS (Global Positioning System) receivers have led to increased application of GIS analysis for plant diversity conservation and management. Increased accessibility and use of the internet has also created a revolution in the sharing of biodiversity, geographical and environmental data.

The Global Biodiversity Information Facility (GBIF), for example, is a platform providing public access to biodiversity data from national museums, herbaria and genebanks worldwide (www.gbif.org). In October 2012, the GBIF contained already more than 75 million geo-referenced plant observations (GBIF 2012). This is a wealth of data that can be used for ecogeographic and biodiversity monitoring studies by scientists and students in different parts of the world. Knowledge on geospatial analyses is required to fully take advantage of the opportunity of increased data availability and analysis. Other important themes are data preparation and cleaning to assure adequate data quality for sound analysis, and how to interpret the results. Across this thesis, the latter points are highlighted several times.

The overall aim of this thesis is to contribute to a better conservation of plant species, with an emphasis on PGR, and reduce the current rate of plant biodiversity loss. More specifically, and in collaboration with local research partners, I will explore the use of geospatial analysis in mapping plant diversity and distribution to support *in situ* conservation and collecting of PGR in the field.

The hypothesis tested in this dissertation is the following: geospatial analysis of plant diversity and distribution can clearly detect geographic inter-specific, intra-specific diversity and distribution patterns, which allows to prioritize plant species populations and geographic distribution areas that should be considered for *in situ* conservation and germplasm collecting. This proposition is tested by (1) mapping diversity and distribution of plant genetic resources in several case studies in collaboration with local research partners; and (2) by asking species' experts their feedback on the usefulness of Environmental Envelope Modelling (EEM) in predicting species distributions. The latter technique is widely used in ecological and biogeographic studies. The results from this study will allow to better understand how useful EEM is to support local practitioners in the implementation of conservation measures.

Discussions on how to measure status and trends of biodiversity are still on-going (see Pereira et al. 2013). The methods and analyses presented in this thesis contribute to the discussions on the development and implementation of regional and global monitoring systems of economic plant diversity.

Related but different biological disciplines, such as resource ecology, plant systematics and PGR conservation may consider plant diversity and distribution analyses from different points of view. These analyses could therefore have different objectives and use of different terminologies. In the case of resource ecology the principle aim may be to understand environmental drivers behind species diversity. The purpose of plant systematics studies is often to define phylogenetic relationships between species, genera and/or clades. Identification of areas for plant germplasm conservation and collecting is the main objective for PGR conservation. Even within a specific discipline, different points of view and theoretical frameworks can lead to the definition of different objectives and the use of different types of analyses/methods.

The plant diversity and distribution analyses presented in this thesis, focus on the application of geospatial analysis for PGR conservation and use. They are carried out from a clear human development perspective because PGR are genetic material for current and future human use. The dissertation builds further on the analyses and geospatial concepts of Guarino (1995), Guarino et al. (2002), Jarvis et al. (2003) and Scheldeman et al. (2007). The standard work of Frankel et al. (1995a,b,c) on conservation of plant biodiversity is often referred to for general genetic concepts in plant diversity conservation.

Building on the work of these researchers and many others, this thesis presents several novel concepts that I hope will further improve the application of spatial analysis for PGR conservation and use. These approaches include (1) the formalization of expert knowledge to validate and improve species distribution models; (2) combination of spatial threat, conservation and distribution analysis to assess the conservation status of economic plant species; and (3) mapping intra-specific diversity of plant species by means of molecular marker characterization applied to sampled specimens and accessions to prioritize areas for conservation.

In the first part of the thesis, concepts and methods of spatial analysis of plant diversity and distribution for *in situ* conservation and germplasm collecting missions are presented. These approaches allow to assess diversity and conservation status of PGR at both species and intra-specific level. This section includes three chapters. The first chapter consists of a general overview of and practical recommendations on the use of GIS and ecogeographic studies for PGR management. In the second chapter, the use of molecular markers in spatial germplasm characterization to optimize *in situ* conservation of PGR is highlighted. Molecular markers have several advantages when one wants to assess and monitor intra-specific plant diversity of economic plants. They are cheap compared to morphological characterization and quantitative genetics, environmental-neutral, and repeatable in time. Initiatives that promote conservation and sustainable use of PGR call for assessments of genetic variation in plant species, with molecular markers as one of the tools prescribed (FAO 2010a; 2011). As a consequence more and more molecular diversity studies are carried out for economically important plant species. At the same time, at species level, still remarkably little is known about the distribution ranges of many plant species, despite the increases in the number of plant observation data (Feely and Silman 2011). There is especially a deficit of plant observation data in subtropical and tropical regions, which harbour high biodiversity compared to boreal and temperate regions, (Nic Lughadha et al. 2005). Therefore, the third chapter discusses the use of EEM to predict geographic distribution ranges of plant species for *in situ* conservation and collecting trips. Scientists and professionals are being asked to evaluate EEM for their conservation and management activities in the field. To analyse their feedback on distribution models in a scientifically rigorous way, a method is presented to formalize expert knowledge on the basis of cultural consensus theory. As mentioned above, the results of this last chapter will be an important input used to test our hypothesis, i.e. that spatial analysis of plant diversity and distribution are useful to support local practitioners in the implementation of conservation measures.

In the second part of this thesis, several Latin American case studies are presented on the diversity, distribution and germplasm conservation of specific plant species of human use and concern. This region harbours areas of exceptionally high vascular plant diversity. They include tropical and subtropical Andes, several parts of Mesoamerica (including Mexico), the Atlantic forest, the Chocó forest in Colombia, and the north-western part of the Amazon (Barthlott et al. 2007).

In addition to being global hotspots of vascular plant diversity, the Andean region and Mesoamerica are also considered to host two centres of plant domestication where agriculture was developed independently (Pickersgill and Heiser 1977; Vavilov 1992a; Pickersgill 2007). Almost 300 cultivated species have been identified as having their origin in South America (Zeven and de Wet 1982). Most of them come from the Andean region. On top of that, at least 225 cultivated species have been identified with their centre of origin in Mesoamerica (Zeven and de Wet 1982). Several globally important food crops originated from Latin America. The list comprises cassava (*Manihot esculenta* L.), common bean (*Phaseolus vulgaris* L.), groundnut (*Arachis hypogaea* L.), maize (*Zea mays* L.), chili pepper species (*Capsicum* spp.), potato (*Solanum tuberosum* L.), squashes and pumpkins (*Cucurbita* spp.), sweet potato (*Ipomoea batatas* L.) and tomato (*Solanum lycopersicum* L.). Other crops are regionally and locally important and remain under-researched. Many of these species were forgotten after discovery of the continent by Columbus. Their potential for wide cultivation has only been re-discovered recently (National Research Council 1989a).

More recently, the Amazon has been identified as a third centre of crop domestication (Clement 1999; Pickersgill 2007; Meyer et al. 2012). This area has been overlooked because of the high rates of crop genetic erosion in this area after 1492 due to (1) rapid human decline due to severe epidemics of European-introduced infectious diseases; (2) presence of only a limited number of archaeological plant remains and; (3) rapid deforestation (Pearsall 1992; Clement 1999). The Amazon may also receive less recognition as being an area of domestication because forest and tree domestication has received traditionally little attention by western archaeological and biological scientists compared to grassland crops such as most cereals (Wiersum 1997). In the Amazon, most plants that were domesticated are tree species (Clement 1999). Amazonian tree crops include Brazil nut (*Bertholletia excelsa* Humb. & Bonpl.), cacao (*Theobroma cacao* L.), cashew (*Anacardium occidentale* L.), cupuaçu (*Theobroma grandiflorum* (Willd. ex Spreng.) K. Schum.), guaraná (*Paullinia cupana* Kunth), ice-cream bean tree (*Inga edulis* Mart.) and peach palm (*Bactris gasipaes* Kunth) (Clement et al. 2010). The tropical lowlands in South America are also believed to be the centre of origin of several root crops, the most important being cassava (Pickersgill 2007; Clement et al. 2010). Eastern North America has been identified as a fourth centre of independent crop domestication in the Americas (Pickersgill 2007). However, only a few crops have been developed in this region in pre-Columbian times including sunflower (*Helianthus annuus* L.) and pumpkin (*Cucurbita pepo* L.) (Pickersgill 2007). Still 113 cultivated plant species have been identified native to the US and Canada (Zeven and de Wet 1982). Several of these species such as highbush blueberry (*Vaccinium corymbosum* L.) and cranberry (*Vaccinium macrocarpon* Aiton) have only been domesticated recently, in the last 500 years after Columbus' discovery of the Americas (Meyer et al. 2012).

Whereas cultivated plants are not necessarily domesticated and can also occur in natural populations, domesticates are per definition cultivated (Clement 1999). In this thesis often wild vs. domesticated and cultivated vs. natural populations are distinguished. This is used as a conceptual framework to explain differences between on-farm conservation and *in situ* conservation of populations in a natural ecosystem. However, in practise, this distinction often cannot easily be made in the field. For tropical tree species, the dichotomy between domesticated and wild species, cultivated and natural populations is not that clear-cut in the field. As mentioned before, many tropical tree species are in incipient phases of domestication and many semi-domesticated plant individuals exist (Clement 1999) Secondly, there are many different types and degrees of human intervention from forest to intensive tree crop cultivation (Wiersum 1997). The latter makes it difficult to delineate cultivated plant populations from natural ones.

Domesticated annual and bi-annual species can be more easily differentiated from their wild progenitors and relatives than most tropical tree species in incipient phases of domestication. Key traits of domestication include modifications in fruit and root morphology (e.g. increase in size) and change in secondary metabolites (e.g. reduction of toxins) to name just a few (Meyer et al. 2012). Nevertheless, in centres of crop origin and diversity (i.e. the areas that should be prioritized for PGR conservation), a plethora of morphotypes can be found that have intermediate values for these traits. This makes distinction between wild botanical varieties and domesticates of annual and bi-annual crops, sometimes hard as well. The continuum from wild to domesticated types is, for example, beautifully demonstrated in Peruvian and Bolivian genebank collections of the chili peppers *Capsicum chinense* and *Capsicum baccatum* from their respective centres of diversity. There are also crops where only the domesticated variety exists. An example is the Andean domesticated chili pepper *Capsicum pubescens*. Intriguingly, its wild progenitor is unknown and probably has gone extinct (Eshbaugh 2012).

Plant domestication started after about 13,000 years ago in different parts of the world including the Andes, Mesoamerica and Eastern North America (Purugganan and Fuller 2009; Meyer et al. 2012). Through time, the intensity of domestication has been dynamic and was related the rise and fall of advanced civilizations (Meyer et al. 2012).

The Andean region went through an intense period of plant domestication about 8,000-10,000 years ago (Meyer et al. 2012). In the Mesoamerican region, two periods with high levels of high plant domestication can be observed, i.e. 8,000-10,000 years ago and 5,000-6,000 years ago (Meyer et al. 2012). In the Amazon, plant domestication seems to have become particularly intensified 2000-500 years (Meyer et al. 2012) albeit some important crops such as cassava were already domesticated at least 6,500 years ago (Mannion 1999). This process stopped abruptly after European invasion (Clement 1999).

The number of studies on domestication processes is growing steadily (Purugganan and Fuller 2009). However, for many plant species there exist significant gaps of information on regions of crop origin, among other missing archaeological and genetic data (Meyer et al. 2012). Molecular genetics can help to specify centres of crop origin, dispersion and diversity (Pickersgill 2007; van Etten and Hijmans 2010). In chapter 6, an approach is presented to map molecular genetic diversity to identify hotspots of diversity for PGR conservation and use.

Central America and the Andes, being the cradle of several crops of worldwide importance, are also home to many CWR (Maxted and Kell 2009). In the first case study of this thesis, the diversity and conservation status of endemic wild potato species in Bolivia is evaluated in a combined threat and ecogeographic analysis. Bolivia is a global CWR hotspot with a high number of wild potato, chili pepper, groundnut, and cassava relatives (Maxted and Kell 2009).

Almost half the territory of the South America region is still covered with forest. This percentage is higher than that of any other region in the world (FAO 2010b) whereas the percentage forest cover in Central America was still 38 % in 2010 (FAO 2010b). At the same time, South American forests are affected by the highest net loss compared to other regions; between the years 2000 and 2010 about four million hectares net loss per year (FAO 2010b). To a lower degree Central America forest cover also continues to decline (FAO 2010b). Among other negative environmental implications, this is of concern for PGR conservation and relevant for germplasm collecting because deforestation is an important driver of genetic erosion (Clement 1999). These forests provide a high number of forest genetic resources (FGR) for timber and non-timber products, such as fruits and resins. For example, more than 1,200 new world fruit species have been described in literature of which many are woody perennials (Bioversity International 2004). In this dissertation, I present the story of genetic resources´ use and conservation for two of these fruit species. The first one, peach palm (*Bactris gasipaes* Kuntz) is the socio-economically most important domesticated palm species native to Latin America (Clement et al. 2010). Palm hearts and fruits of this species are much-appreciated in several Neotropical areas because of their nutritional value and taste whereas they provide a good crop alternative for small-scale farmers (Graefe et al. 2013). A general overview will be given of the species’ genetic diversity, distribution and genetic resources management. The second fruit species, cherimoya (*Annona cherimola*

Mill.) has been praised for its excellent taste and smell (National Research Council 1989b). The fruit provides national and international market opportunities for Andean small-scale farmers (Vanhove and Van Damme 2009) and is also widely cultivated in Mediterranean climate regions around the world, principally in Spain, United States and Chile (van Zonneveld et al. 2012). Cherimoya will be used as a model species to demonstrate how characterization with molecular markers can be used in combination with spatial analysis approaches to identify centres of genetic crop diversity and support conservation and use of PGR.

If not mentioned otherwise, the Germplasm Resources Information Network (GRIN) Taxonomy plant species nomenclature is followed. However, in the specific case of wild potato relatives (*Solanum* spp.) endemic to Bolivia taxonomy follows Spooner and Salas (2006). This nomenclature is generally applied in global databases and the Bolivian potato genebank collection.

The studies are all presented under the assumption that clear and practical spatial analysis results can be used by local practitioners to develop more specific and efficient *in situ* conservation and germplasm collecting strategies, which eventually will lead to improved plant genetic resources conservation. Therefore, to enhance local use of these analytical tools, a training manual has been developed in parallel to this thesis for scientists, professionals and students who work in biodiversity conservation and are interested to apply spatial methods to analyse geographic patterns of plant diversity and distribution (Scheldeman and van Zonneveld 2010). This manual includes examples of basic analyses that are developed in much more detail in this dissertation and is freely accessible online (http://www.gbif.org/orc/?doc_id=2917&l=en). In the general discussion of this thesis, I will summarize the conclusions and results of the concepts and application, and further discuss how the results of these analyses can be used on the ground.

Concepts

1. Ecogeographic analyses and GIS tools: general overview and practical recommendations for germplasm collecting and *in situ* conservation planning[i]

Maarten van Zonneveld[1,2], Evert Thomas[2], Gea Galluzzi[2] and Xavier Scheldeman[2]

[1] Ghent University, Faculty of Bioscience Engineering, Coupure links 653, 9000 Gent, Belgium
[2] Bioversity International, Regional Office for the Americas, Apartado Aereo 6713, Cali, Colombia

Abstract

Ecogeographic studies provide critical information on plant genetic resources (PGR) to assess their current conservation status and prioritize areas for conservation. They have also proven to be useful for planning for effective genebank management, such as the definition of core collections and identification of collection gaps. In this update, we give an overview of relevant techniques and advances in ecogeographic studies of PGR that can be used to analyse biodiversity data based on field-collected data and to target further collecting. We commence with providing some general recommendations that are important when setting up new research projects that are aimed at assessing the conservation status of PGR and/or monitoring trends in (agricultural) biodiversity on the basis of ecogeographic analysis. A brief introduction to commonly used methods and techniques for the analysis of inter- and intra-specific diversity is provided. The latter include multivariate methods such as clustering and ordination. We also elaborate on mapping of (agricultural) biodiversity data and emphasize the importance of ensuring good data quality. Furthermore, we provide a synopsis of methods for distribution modelling and present an overview of useful open-access and commercial statistical and GIS packages.

---

[i] Adapted from: van Zonneveld M, Thomas E, Galluzzi G, Scheldeman X (2011) Chapter 15/16: Mapping the ecogeographic distribution of biodiversity and GIS tools for plant germplasm collectors. In: Guarino L, Ramanatha Rao V, Goldberg E (eds) *Collecting Plant Genetic Diversity: Technical Guidelines - 2011 Update*. Bioversity International, Rome, Italy.
http://cropgenebank.sgrp.cgiar.org/index.php?option=com_content&view=article&id=662

1.1 Introduction

Ecogeographic studies refer to the process of collecting, characterizing, systemizing and analysing different kinds of data pertaining to target taxa within a defined region (Maxted et al. 1995). These kinds of studies are important for the formulation and implementation of more targeted and, hence, more effective conservation strategies for plant genetic resources (PGR) (Guarino et al. 2005).

Taxonomic, morphological and genetic data can provide critical information about the diversity present in specific geographic areas, which, in turn, can be used for various purposes, such as the assessment of the current conservation status of PGR and to prioritize areas for *in situ* conservation. At the *ex situ* level, combining climate and other ecological information of an accession's collection site – from its passport data – with corresponding morphological and/or molecular characterization data has also proven useful for effective genebank management (e.g. definition of core collections, identification of collection gaps, etc.). Geographic information systems (GIS) are useful tools for this type of analysis (Guarino et al. 2002). They allow to perform complex analyses. Through GIS it is possible to visualize results in clear maps. This facilitates decision making by relevant authorities and encourages the development and implementation of conservation policies (Jarvis et al. 2010). GIS analysis is carried out on the basis of a coordinate system in which each point location at earth has a unique *x* and *y* value. Hence, it is important to access good-quality georeferenced biodiversity data in ecogeographic studies.

1.2 Preliminary data handling

The following three paragraphs present several key recommendations on how to initiate an ecogeographic survey for PGR, following Guarino et al. (2005). Any such study should start with a workplan that clearly states the objectives and the methodological design, including a sound strategy for data collection. Taxonomical experts should be identified who can provide key information about the target taxa and validate the results/products obtained from ecogeographic analyses and research, such as distribution maps and the results of collection gap analysis (Ramirez-Villegas et al. 2010). When available, it can be extremely useful to involve networks of taxonomical experts in such studies. Species experts from the Latin American Forest Genetic Resources Network (LAFORGEN) have, for example, provided basic information about reproductive behaviour (breeding systems, pollination and seed dispersal systems) of prioritized tree species in the MAPFORGEN project (www.mapforgen.org). MAPFORGEN is a collaborative platform of researchers and institutions that presents information to support the conservation of 100 socio-economically important woody perennial species native to Latin America and the Caribbean.

Given the continuous changes in taxonomical classification of plants (APG III 2009), it is of utmost importance to determine upfront the taxonomical boundaries and nomenclature that will be used. In this respect, the online database of the US Germplasm Resources Information Network (GRIN) (www.ars-grin.gov/cgi-bin/npgs/html/index.pl) provides a useful reference for economic plants. Nevertheless, it is strongly advisable to consult other databases such as the Plant List (www.theplantlist.org) or the International Plant Names Index (IPNI) (www.ipni.org) as well as to refer to other data sources such as experts, monographs and Floras when defining the nomenclature to follow.

The geographical extent and boundaries of the target region depend on the objectives of the study. For example, a study focusing on assessing the status of PGR for strengthening national conservation programmes will be limited to the country's national territory. In most other cases, since the occurrence of cultivated and wild taxa does not follow political boundaries, the target region of ecogeographic studies will be defined based on available knowledge about the distribution and diversity of taxa, compiled from literature reviews (e.g. Zeven and De Wet 1982) and consultation with experts from national or international agricultural research centres.

Data collection

Before starting actual collection of field data, the preparation of a clear list of descriptors for passport data is recommended. Data standards for multicrop descriptors have been developed to standardize passport data, morphological characterization and evaluation. These standards make the resulting information comparable across herbarium and germplasm samples (Alercia et al. 2012). This facilitates the use of data from different sources in comparative analyses. In a similar manner, in order to enable comparison of molecular characterization of crop species, minimum standard sets of markers have been suggested (Van Damme et al. 2011).

Original field notes should be saved carefully and adequately backed up to allow for cross-checking of data at a later stage. A backup should also be made of the original data files stored in a notebook or Global Positioning System (GPS) receiver. Field data can be integrated with additional data retrieved from online portals comprising data from genebanks and herbaria, contributing to more comprehensive analyses on the distribution and conservation of PGR (see table 1.1 for an overview).

Table 1.1 Online PGR documentation systems and portals for sharing biodiversity data.

Portal	Data type	Website
Germplasm Resources Information Network (GRIN), National Plant Germplasm System (NPGS)	Passport, characterization and taxonomic information of PGR conserved by the United States Department of Agriculture (USDA)	www.ars-grin.gov/npgs/index.html
System-wide Information Network for Genetic Resources (SINGER)	Passport data of the PGR conserved by the Consultative Group on International Agricultural Research (CGIAR) Centres	http://singer.cgiar.org
EURISCO	Access to all *ex situ* PGR information in Europe	http://eurisco.ecpgr.org
Genesys	Passport, characterization and evaluation data for the 22 most important crops, from CGIAR Centres, EURISCO and GRIN	www.genesys-pgr.org
Global Biodiversity Information Facility (GBIF)	Passport data from herbaria and genebanks from all around the world	www.gbif.org
SpeciesLink	Passport data from the Brazilian herbarium information system	http://splink.cria.org.br/index?&setlang=en
JSTOR Plant Sciences	Taxonomic information and historic herbarium samples	www.plants.jstor.org
Botanical Research and Herbarium Management System (BRAHMS)	Instructions for mapping species distribution summaries and diversity indices	http://dps.plants.ox.ac.uk/bol

Recording geographical data is normally done directly in the field by assigning geographical coordinates through the use of a GPS receiver. The geographic coordinate system in GPS receivers can usually be adjusted according to the user's preferences. Two commonly used coordinate systems are longitude/latitude and Universal Transverse Mercator (UTM). Longitude/latitude is preferred in large-scale studies, such as for mapping the distribution range of taxa that occur across different countries. The longitude/latitude coordinate system in combination with the World Geodetic System (WGS) 1984 is recommended in data standards for multicrop descriptors (Alercia et al. 2012). It is the coordinate system used in many freely available spatial datasets (see table 1.2 for an overview). This makes it the preferred option in combination with WGS 1984 for combining different spatial datasets.

For studies at lower administrative units (e.g. province, department, state), UTM may be preferred because of the low distortion at this scale and the ease in calculating geographic distances. To be able to carry out GIS analysis with the collected data, longitude/latitude coordinates should be in decimal degrees. If longitude/latitude coordinates of collection sites were listed in degrees, minutes and seconds, a special formula can be applied to convert these coordinates into decimal degrees (see chapter 2 of Scheldeman and van Zonneveld 2010).

Table 1.2 Some spatial data sources and tools.

Climate
• Interpolated climate surfaces for the globe up to one km resolution: WorldClim (www.worldclim.org) • Downscaled layers from future climate models (GCMs): Climate Change Agriculture and Food Security (CCAFS) (www.ccafs-climate.org) • Reconstructed paleoclimates: US National Oceanic and Atmospheric Administration (NOAA) (www.ncdc.noaa.gov/paleo/paleo.html)
Topography
• Elevation, watershed and related variables for the globe at one km resolution: US Geological Survey (USGS) (http://eros.usgs.gov) • High-quality elevation data for large portions of the tropics and other areas of the developing world: SRTM 90 m Elevation Data (http://srtm.csi.cgiar.org)
Remote sensing (satellite)
• Various land-cover datasets: Global Land Cover Facility (GLCF) (http://glcf.umiacs.umd.edu/data) • Various atmospheric and land products from the MODIS instrument: National Aeronautics and Space Administration (NASA) (http://modis.gsfc.nasa.gov/data)
Soils
• Harmonized World Soil Database (www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML)
Other spatial data
• Relevant links and data at DIVA-GIS website (country level, global level, global climate, species occurrence); near global 90-meter resolution elevation data, high-resolution satellite images (LandSat) (www.diva-gis.org/Data) • Spatial database of the world's administrative areas (or administrative boundaries): Global Administrative Areas (GADM) (www.gadm.org) • Database with eight million place names with geographical coordinates: GeoNames (www.geonames.org)

Since various identification codes may be used in the different steps of collecting, characterizing and evaluating germplasm material (e.g. collector code, field code, collection code), it is essential to clearly define a unique identification code to be applied to each accession throughout the entire study. This will ensure consistent and unequivocal correspondence between each accession and the complexity of its passport, characterization and evaluation data. The latter approach is key to getting trustworthy georeferenced taxonomic, phenotypic or genetic diversity data for ecogeographic studies.

For example, the definition and consistent use of unique identification codes of the Peruvian chili pepper collection maintained by the Peruvian agricultural research institute (INIA) have been essential to link biochemical characterization carried out in a laboratory to the correct taxonomy determined in an experimental field and the right passport data recorded during germplasm collecting (Meckelmann et al. 2013). The addition of new codes should be considered with care. More codes may lead to confusion and increase the likelihood of making errors in the documentation system. This affects the reliability of the data and reducing the possibility of effectively conserving and using collected and characterized germplasm.

Diversity analyses

Ecogeographic studies related to the conservation and use of PGR are mostly focused at the species or gene levels of plant diversity. At species level, the observed unit of diversity is the species on the basis of taxonomic identification, measured mostly as presence or absence in a certain location (species richness). Other parameters of species diversity are evenness and abundance (Magurran 1988). Studies at the gene level can be either inter-specific (e.g. phylogenetic studies within a gene pool or clade) and/or intra-specific (i.e. to understand genetic variation between plant individuals of the same species or within and between populations of plant species).

For the purpose of measuring genetic variation, the chosen units of diversity may be phenotypic traits (the products of a gene or its expression) or, more directly, variation in sequences of neutral or functional portions of DNA or RNA, measured with the assistance of molecular markers (e.g. SSRs, SNPs, DArT, AFLPs; see De Vicente and Fulton [2004] and Kumar et al. [2009] for an overview of various widely applied molecular markers).

Richness in species or in the number of alternating DNA sequences in specific parts of a plant species genome (e.g. allelic richness) are straightforward measures of diversity and are commonly used for prioritizing conservation areas of either plant communities – based on number and uniqueness of observed species (Gotelli and Colwell 2001) – or within-species populations identified through molecular marker characterization (Frankel et al. 1995b; Petit et al. 1998). However, richness is sensitive to sampling bias – the situation where an uneven number of observations or collections has been made across the sampling units included in an ecogeographic study (some units will contain more observations than others). The rarefaction methodology allows correcting such sampling bias by recalculating richness on the basis of an equal, user-defined number of observations per sampling unit (Petit et al. 1998; Gotelli and Colwell 2001). Another possibility is re-sampling without replacement to a minimum sample size (Leberg 2002).

In studies of genetic diversity based on molecular markers, the number of locally common alleles is an important indicator for prioritizing populations of wild and domesticated plant species for *in situ* conservation. These alleles occur in relatively high frequency over a limited area and can evidence long histories of local adaptation to specific environments and areas that historically have been isolated (Frankel et al. 1995b). Locally common alleles can be identified by statistical programmes for genetic data such as GenAlEx (see table 1.3), which identifies alleles with a frequency higher than 5 % in a local population and occurring in less than 25 % of all populations as locally common alleles (Peakall and Smouse 2006). Another way to detect locally common alleles is with the help of GIS, by identifying those alleles that occur at relatively high frequencies within a given maximum distance (see chapter 5 of Scheldeman and van Zonneveld 2010). In the next chapter, we discuss the application of molecular marker characterization in geospatial analysis in more detail.

Distance parameters

In diversity analysis, ecological and genetic distances are statistics of central importance that allow investigating the existence of structure and patterns in biodiversity data (*beta* diversity). This, in turn, is essential for prioritization exercises for *in situ* conservation (Gallo et al. 2009; Petit et al. 1998; van Zonneveld et al. 2012), as well as for germplasm management and use, such as in the establishment of core and reserve collections (Frankel et al. 1995a). Ecological distances can be used to calculate how divergent different sampling units are based on their species or varietal composition. On the other hand, genetic distances are typically used to calculate how divergent within-species individuals or populations are, based on morphological trait or allelic composition. Genetic distances can also be used in phylogenetic studies to classify the evolutionary position of species. Multivariate techniques such as clustering and ordination allow the ordering of units of diversity, such as sampling units, species, plant individuals (within species), on the basis of the ecological or genetic distances between them.

Several open-access analysis packages can be used for carrying out diversity analyses, including the calculation of distance parameters, clustering and/or ordination analyses. Some commonly used programmes for ecological and genetic diversity, and structure analyses are listed in table 1.3. Additional software for specific genetic analyses is listed in Appendix A of Lowe et al. (2004) and in Excoffier and Heckel (2006).

There is a wide variety of different distance statistics that can be employed, each with different properties. Some distance measures, such as the Euclidean distance, are used for calculating both ecological and genetic distances, whereas other measures are generally used for either one of them. Other popular ecological distances include Bray-Curtis, Kulczynski, Hellinger and Chi-square distances (Kindt and Coe 2005). Since distance measure is the form for subsequent multivariate techniques (e.g. clustering, ordination) and will thus affect the results of this type of analyses, it is important to select an appropriate distance statistic. A desirable characteristic of any ecological distance parameter is that it assigns the same maximum distance to all pairs of sites that do not have any species in common [e.g. Bray-Curtis and Kulczynski distances (Kindt and Coe 2005)]. For other features of different ecological distance parameters and how to test them, refer to Kindt and Coe (2005).

Table 1.3 Open-access applications for biodiversity and genetic analyses.

Software	Properties and applications	Source
Biodiversity.R	A single software environment for performing nearly all types of biodiversity analysis; Operates in statistical programme R	Kindt and Coe 2005 http://cran.r-project.org
Vegan	Ordination methods, diversity analysis and other functions for community and vegetation ecologists; Operates in statistical programme R	Kindt and Coe 2005 http://cran.r-project.org/web/packages/vegan/vignettes/intro-vegan.pdf
Biodiversity-Pro	Alpha and beta diversity analysis, multivariate statistics	McAleece et al. 1997 http://gcmd.nasa.gov/records/NHML_Biopro.html
EcoSim	Null model analysis in community ecology	Gotelli and Entsminger 2004 http://garyentsminger.com/ecosim/index.htm
PAST	Developed for palaeontology, but offering vast possibilities for (multivariate) biodiversity analysis	Hammer et al. 2001 http://folk.uio.no/ohammer/past
GenStat Discovery	Free version of statistical programme GenStat	www.vsni.co.uk/software/genstat-discovery
Adegenet	Population genetics, including clustering based on Bayesian Information criterion, Discriminant Analysis of Principal Components and spatial Principal Components Analysis; Operates in statistical programme R	Jombart 2008 http://adegenet.r-forge.r-project.org
Structure	Free software package for using multi-locus genotype data to investigate population structure.	Pritchard et al. 2000 http://pritch.bsd.uchicago.edu/structure.html

GenAlEx	User-friendly cross-platform package for population genetic analysis Runs within Excel	Peakall and Smouse 2006 www.anu.edu.au/BoZo/GenAlEx

The choice of genetic distance measures largely depends on the type of data (phenotypic, dominant or co-dominant molecular marker characterization data) and on whether distance is calculated between individuals or between groups of individuals. A guide to which measures of genetic distances may be most appropriate for different situations is provided by Lowe et al. (2004) and De Vicente et al. (2004b). Popular genetic distance parameters include Nei's standard genetic distance, Arc distance or Manhattan distance for quantifying distances between populations, and Tanimoto or Jaccard distance for quantifying distances between individuals (Geburek and Turok 2005).

A series of distance parameters can be used when estimating the variation in phenotypic traits between individuals of the same species. This applies to data analyses from so-called 'common-garden' experiments (e.g. Willemen et al. 2007). In such experiments, plant material collected in different sites is planted in field trials under a common environment, in order to reduce the variance by the environmental effect in the expression of phenotypic traits. The Gower distance can be used when a dataset contains data of both nominal (e.g. orthogonal and categorical) and continuous morphological variables (Grum and Atieno 2007; Willemen et al. 2007). The Ward-MLM distance (Franco et al. 2010) is useful for combining phenotypic and molecular characterization data for clustering or ordination. In light of the different properties of the different genetic distance statistics, it is important to note that care must be taken when comparing different studies that use different distance parameters (Finkeldey 2005).

Distance measures can also be used to test the hypothesis that individuals that are geographically located far away from each other are also genetically more distant according to the isolation-by-distance model developed by Wright (1943). To do this, the Mantel correlation value is often used to calculate between pairwise geographical and genetic distances. Mantel tests can be carried out in packages via Adegenet or GenAlEx (see table 1.3). Other types of distances can be compared with genetic distances through Mantel tests as well, such as environmental climate and soil distances in order to examine whether individuals from different ecological zones are also genetically more distinct (Kozak et al. 2008). In GIS programmes, environmental data (climate, topography, soils) for each collection site can be easily extracted from freely available spatial data maps and exported to a spread sheet for further statistical analysis (Scheldeman and van Zonneveld 2010). Table 1.2 provides an overview of important sources and tools for spatial data analyses.

Clustering

Clustering refers to methods that draw on the distance parameters discussed above for assigning units of diversity into groups or clusters whose members show a certain level of similarity for the measured characteristics. Units of diversity can refer to sampling units, species, within-species individuals or specific populations. Many hierarchical and non-hierarchical clustering methods exist and it is practically impossible to choose a 'best' method among them because of their heuristic nature. The robustness of clustering is limited because the outcomes can change substantially depending on different combinations of distance parameters and clustering methods. Therefore, this type of analysis is specifically useful for exploring variation within collected data, and not for definitive multivariate analyses (Kindt and Coe 2005). Whereas a markedly discontinuous structure in data will likely be detected by almost any method, a more gradual or continuous makeup of inter- and intra-specific plant diversity will be more difficult to detect by cluster analyses (Jongman et al. 1995). In these cases, ordination methods are more appropriate than clustering methods (Kindt and Coe 2005). It is possible to evaluate the clustering performance of a distance statistic by calculating cophenetic correlation. The latter compares the distances between observation points calculated by a given distance parameter with the corresponding distances between these points in the cluster diagram [for further information see Kindt and Coe (2005)].

Grum and Atieno (2007) provide a user-friendly introduction to clustering with continuous and nominal variables in the free statistical programme R (R Development Core Team 2010). A frequently used programme to assign plant individuals to genetic clusters on the basis of molecular characterization data is Structure (Pritchard et al. 2000), which uses a Bayesian approach to determine the probabilities of plant individuals belonging to each cluster from a predefined number of clusters. These clusters can also be geographically visualized in GIS (Vigouroux et al. 2008).

Ordination

The basic aim of ordination is to represent observations (e.g. the occurrence of different species across a climate gradient, or the allelic composition of plant individuals from a specific gene pool) and sampling units (e.g. different plots in which species compositions are determined, or sample tissues from different individuals that are used for determining allelic composition) in a two-dimensional space in such a way that points that are close together are considered more similar than points that are further apart. Ordination allows simultaneous representation of observations and geographic sampling units in the same plane. Observations of species or plant individuals (within a species) that are plotted close together have a higher likelihood to occur in sampling units with more similar characteristics (e.g. because they share the same environmental niche, or morphological or molecular characteristics) as compared to points that are plotted further apart. Likewise, points representing geographic sampling units that are close together correspond to sampling units that are similar in species, morphological trait or allelic composition, whereas points that are far apart correspond to samples that are dissimilar in these elements. This combined visualization allows one to relate patterns in observations with underlying explanatory patterns in relative geographic sampling units (for instance, between-species similarity and similarity between plots where these species were observed).

Two general approaches are used in ordination. In direct (or constrained) gradient analysis, direct relationships are sought between (1) occurrence and/or abundance of species, varieties or alleles; and (2) specifically measured (environmental) variables that characterize the geographic sampling units in which these species, varieties or alleles were observed. Observations and geographic sampling units are arranged in a virtual space along axes that are linear combinations of these explanatory variables (e.g. environmental variables). Thereupon, the predictive power of each of the respective variables is determined (Höft et al. 1999). By contrast, indirect (or unconstrained) gradient analysis focuses entirely on observations and allows maximum explanation of variation without the restriction of explanatory variables (Jongman et al. 1995). This type of analysis is particularly useful when there is no clear foreknowledge about variables that might explain variation between the observations.

Most types of direct and indirect gradient analysis can be divided into two main types of ordination techniques: those that are related to (1) a linear (monotonic) response model in which the abundance of any observational unit (such as species or within-species plant individuals) either increases or decreases with the value of each of the explanatory variables (e.g. Principal Components Analysis [PCA] and Redundancy Analysis [RDA]); and (2) a unimodal response model, where any observational unit occurs within a limited range of the explanatory variables (e.g. Correspondence Analysis [CA] and Canonical Correspondence Analysis [CCA]) (Jongman et al. 1995). Given that the unimodal distribution is more common in nature than a linear distribution, it might be more advantageous to use unimodal over linear response models (Kindt and Coe 2005). According to Jongman et al. (1995), it is advisable to start analysing biodiversity data by using unimodal models (CA, Detrended Correspondence Analysis [DCA] or CCA) and to decide afterwards and based on these first results whether one could simplify the model to a monotonic one. Non-Metric Multidimensional Scaling (NMDS) is an additional method for indirect gradient analysis that differs in various ways from nearly all other ordination techniques. It can handle non-linear species responses of any shape and allows the use of any distance parameter (Holland 2008). Table 1.4 provides a summary of the different ordination techniques that have been discussed above.

Table 1.4 Ordination techniques.

	Unconstrained or indirect gradient analysis	Constrained or direct gradient analysis	Distance measure
Unimodal response model	• Correspondence Analysis (CA)	• Canonical Correspondence Analysis (CCA)	Chi-square distance
	• Detrended Correspondence Analysis (DCA)	• Detrended Canonical Correspondence Analysis (DCCA)	Chi-square distance
Monotonic or linear response model	• Principal Components Analysis (PCA)	• Redundancy Analysis (RDA)	Euclidean distance
	• Principal Coordinates Analysis (PCoA) = metric multidimensional scaling	• Distance-based Redundancy Analysis (db-RDA)	Any distance
		• Canonical Analysis of Principal Coordinates (CAP)	Any distance
Non-linear response of any shape	• Non-metric multidimensional scaling		Any distance

1.3 Mapping plant diversity data

Data quality control

In mapping the ecogeographic distribution of the target taxa, it is crucial for the data to be of high quality and precise (i.e. to contain a minimum number of errors at a specified scale of study). Therefore, it is very important to check the quality of the data before they are used in analysis. During field collection, it is recommended that detailed passport information is noted down in a field book. Then this original information should be carefully saved to enable to track back any errors that might emerge during data analyses. Chapman (2005) and chapter 4 of Scheldeman and van Zonneveld (2010) explain several ways to check the quality of georeferenced data, including verification of consistency between data on (1) the administrative unit (country, province, department) mentioned in the passport data of a collection or observational record as it was registered in the field; and (2) the administrative unit in which it is mapped in a GIS programme.

Another way to identify potentially erroneous points is to carry out an outlier analysis, which identifies georeferenced taxa data that are located in atypical climates compared to the climatic niche in which records of the taxa normally occur (Scheldeman and van Zonneveld 2010). This type of data can be erroneous due to incorrect coordinates or taxonomic misidentification. However, they might also effectively represent individuals that were observed at the marginal ends of a taxon's distribution range. When this is indeed the case, they could contain valuable traits for adaptation to atypical site conditions. Yet another possibility is that areas with a distinct climate, where outliers are located, have been under-sampled in comparison to other areas. If this is the case, these areas should be considered for further collection. For these reasons, when possible, it is recommended that the field book containing original passport information of a record in an atypical climate be consulted (or else to contact the collector in case the data came from a third party) to find out whether the record is an error. If the outlier appears not to be an error, it can be useful to further evaluate the properties of the plant individuals located in the outlier location based on molecular or phenotypic characterization. If plant individuals possess properties of human interest, it can be worth considering further exploration of the surrounding areas for other plant individuals/populations with similarly interesting traits.

One should also bear in mind that in many cases, data originating from herbaria and genebanks (e.g. freely available from GBIF) were not generated for the purpose of biogeographic studies. They are often the result of ad hoc collecting or non-systematic and uneven sampling efforts (Chapman 2005). Frequently, specimens/accessions have been collected mostly or exclusively from areas that are easily accessible or where a taxon is known to occur, thus negatively affecting the representativeness of the data (Hijmans et al. 2000). Such sampling bias can later be corrected – although only to a certain extent – with methods such as rarefaction and distribution modelling [see Scheldeman and van Zonneveld (2010) for further details]. The best way to prevent sampling bias is, of course, by establishing a sound strategy for data collection, although it should be acknowledged that this is not always possible.

Georeferencing

Georeferencing, which assigns geographical coordinates to collection records or observation data missing such coordinates, can substantially increase the number of sound observation records of the target taxa and thus improve the quality of subsequent ecogeographic studies. Accession and specimen passport/ label data from collections such as herbaria, which do not include geographical coordinates but do include precise information about the locality where the specimen was collected or observed, can be georeferenced either through using gazetteers that can be downloaded from the DIVA-GIS website (see table 1.2) or automated, online gazetteers such as GeoNames (www.geonames.org). Google Earth can be useful for georeferencing observation data as well, especially those that are taken at a specific distance along the road between two localities.

Plant diversity, distribution and conservation

The number and frequency of occurrence of species, varieties or alleles in distinct geographic sampling units within a study area (*alpha* diversity) are the principal subjects of the geospatial analysis of diversity. They allow to prioritize areas for *in situ* conservation and PGR collecting missions. Sampling units may refer to previously identified ecologically specific sites, administrative units or grid cells of any chosen size. In many cases, species distribution is mapped on the basis of observed species presence in the cells of a grid that covers the study area. At a national or continental level, this grid size may be as large as 50 x 50 km, as used in the Atlas Florae Europaeae (2011), or 100 x 100 km (about one degree) (Scheldeman et al. 2007). In this respect, chapter 5 of Scheldeman and van Zonneveld (2010) provides working examples to practice how to map species and allelic richness in grid cells with a point-to-grid analysis in DIVA-GIS.

The advantage of using grid cells is that they allow to compare diversity and divergence between sampling units of similar geographical size over the full extent of the study area. DIVA-GIS and other GIS programmes –among those reviewed in Steiniger and Bocher (2008) – can be used to carry out grid-based diversity analysis (see table 1.5 for open-access and commercial packages). They have been applied in several studies to assess the distribution and conservation status of crop gene pools (e.g. Hijmans and Spooner 2001; Jarvis et al. 2003; Scheldeman et al. 2007). Other ways to map distribution and richness are by means of circular area (Hijmans and Spooner 2001) or circular neighbourhood (Hijmans et al. 2005a; Scheldeman and van Zonneveld 2010). The latter is a re-sampling approach to calculate the values of diversity parameter over all species or allele records found within a specific radius around a grid cell. In chapter 6 of this work a specific case study with the Neotropical fruit tree species cherimoya is presented to map genetic diversity using circular neighbourhood methods,

A number of methods have been developed to optimize the number of conservation areas based on the number of species, varieties or alleles in different units and how they complement each other. DIVA-GIS also include a reserve selection algorithm, developed by Rebelo and Siegfried (1992). This algorithm calculates the minimum number of areas (grid cells) necessary to conserve a given number of species, varieties or alleles of the gene pool under study (Hijmans et al. 2005a). It ranks grid cells that should be given priority for conservation in the following order: first priority is given to the grid cell with the highest alpha diversity; subsequent priority is given to those grid cells that best complement the initial ones because they contain the highest number of new species, varieties or alleles that were not found in the previously selected grid cells (*beta* diversity). Chapter 5 of Scheldeman and van Zonneveld (2010) explains how to carry out such a reserve selection.

Environmental envelope modelling

For most plant species, including many crop wild relatives and socio-economically important tree species, only a limited amount of information on their distribution is currently available (Nic Lughadha et al. 2005). Niche modelling or Environmental Envelope Modelling (EEM) is considered a useful tool for overcoming the lack of concrete information on the distribution of a species (Guarino et al. 2002; Araújo and Peterson 2012). It aims to distinguish between zones where the species could potentially occur (i.e. areas with similar environmental conditions to the defined ecological niche) and areas where the species is likely to be absent because the environment is there different from the ecological niche. The ecological niche of a species is defined on the basis of statistical (empirical) relations between species occurrence and absence sites, and corresponding environmental factors (Araújo and Peterson 2012).

Distribution models can be used for different purposes. This depends on the status of knowledge about a species distribution. First, distribution models can be used to collect data, i.e. to predict on the basis of limited, incomplete knowledge of species occurrence, promising areas for germplasm collecting and *in situ* PGR conservation. Secondly, EEM is used to better understand species-environment relationships. The latter requires a big amount of already collected data on species occurrence.

GIS are very useful in this respect because they allow extraction of information from environmental data layers related to sites where a species has been observed, as well as to sites where it is known to be absent, and allows visualizing and editing the outcomes of the model on a map. Environmental data layers in distribution modelling can be derived from datasets like those listed in table 1.2. Depending on the modelling programme used, they can consist of only continuous variables, such as climate data derived from WorldClim (Hijmans et al. 2005b), or else also include nominal variables, such as maps of vegetation or soil type. While EEM is traditionally used to predict the distribution of species, it could potentially also be applied for predicting intra-specific units of diversity, such as ecotypes or clusters defined on the basis of morphological or molecular characterization. In chapter 4 of this thesis an example is presented of an ecogeographic classification of putative wild potato ecotypes according to the climate zone where samples of these species were collected.

The collection of absence records is a challenge because the reasons for absence are not always clear. Absence might either be due to ecological characteristics, human disturbance or simply because species presence was overlooked during an inventory or collection. Therefore, EEM often uses presence records only (Pearce and Boyce 2006). Presence records can be derived from herbarium specimens, genebank accessions or vegetation/plant species inventories, which been made increasingly available by herbaria and genebanks through online portals such as GBIF (see table 1.1).

In addition to yielding a better understanding of the potential distribution range of a species, distribution models have also been used in gap analyses to prioritize areas for germplasm collecting (Jarvis et al. 2005; Scheldeman et al. 2007). In this respect, a gap refers to a location where a distribution model predicts the potential occurrence of a target taxon, but where specimens and/or germplasm of the taxon have never been collected before. Ramirez-Villegas et al. (2010) present a method based on the identification of sampling, geographic and environmental gaps to prioritize among taxa. Chapter 6 of Scheldeman and van Zonneveld (2010) explains how to carry out a gap analysis with the use of the EEM software Maxent and the GIS tool DIVA-GIS. An important source of guidance is the GapAnalysis portal (http://gisweb.ciat.cgiar.org/GapAnalysis) with its methods for crops and crop wild relatives.

Table 1.5 GIS packages.

Open-source desktop GIS	Properties	Source
DIVA-GIS	Biodiversity analysis, species distribution mapping, etc. Also provides free spatial data for the whole world	www.diva-gis.org
GRASS (Geographic Resources Analysis Support System)	Analysis and scientific visualization, cartography, simulation	http://grass.itc.it/intro
QGIS (Quantum GIS)	Viewing, GRASS-Graphical User Interface	http://qgis.org
uDig (User-friendly Desktop Internet GIS)	Viewing, editing, analysis	http://udig.refractions.net
SAGA (System for Automated Geoscientific Analyses)	Analysis, modelling, scientific visualization	www.saga-gis.org
ILWIS (Integrated Land and Water Information System)	Analysis, integrating image, vector and thematic data	www.itc.nl/Pub/Home/Research/Research_output/ILWIS_-_Remote_Sensing_and_GIS_software.html
OpenJUMP 2002/03	Editing, analysis JUMP Family (Java Unified Mapping Platform)	www.openjump.org

Commercial GIS	Properties	Source
Esri	Products include ArcView 3.x, ArcGIS, ArcSDE, ArcIMS, ArcWeb services and ArcGIS Server.	www.esri.com
Autodesk	Products include Map 3D, Topobase, MapGuide and other products that interface with its flagship AutoCAD software package	http://students.autodesk.com/?nd=download_center&c_key=31305F5F416D657269636173&
Bentley Systems	Include Bentley Map, Bentley Map View and other products that interface with its flagship MicroStation software package	www.bentley.com/en-US/Products/Bentley+Map
ERDAS IMAGINE	Products by ERDAS Inc, include ERDAS ER Mapper, ERDAS ECW JPEG2000 SDK	www.erdas.com/products/ERDASIMAGINE/ERDASIMAGINE/Details.aspx
Intergraph	Products include G/Technology, GeoMedia, GeoMedia Professional, GeoMedia WebMap, and add-on products for industry sectors, as well as photogrammetry	www.intergraph.com
MapInfo	Products by Pitney Bowes, include MapInfo Professional and MapXtreme	www.pbinsight.com/welcome/ten-five/index3.php
Smallworld and Spatial Eye	Purchased by General Electric and used primarily by public utilities	http://site.ge-energy.com/prod_serv/products/gis_software_2010/en/index.htm www.spatial-eye.com/Engels/Spatial-Workshop-features/Direct-access-to-data-in-smallworld-GIS/page.aspx/49

Another application of EEM is to examine the impact of climate change on the distribution of plant species of interest and socio-economic importance, such as crop wild relatives (Jarvis et al. 2008) or timber tree species (Saénz-Romero et al. 2006; van Zonneveld et al. 2009a).

It is important to note that EEM can be used to better understand species distribution and to help prioritize areas for germplasm collection only when some information about a species is already available. There is no standard in terms of the minimum number of observation points required, as this will often relate to the nature of the species: for rare species or species with a restricted niche, only a small number of presence records may be sufficient, while for species with a broad niche and extensive distribution range, a higher total number of records is desirable.

Although it is difficult to provide strict guidelines on the minimum number of presence records that are needed for credible distribution modelling, a number of illustrative examples exist:

- Scheldeman et al. (2007) used a minimum of 10 points for rare *Vasconcellea* species with a known, restricted distribution;
- The MAPFORGEN project (www.mapforgen.org), which evaluates the natural distribution of 100 species native to Latin America, used a minimum number of 20 species presence records;
- van Zonneveld et al. (2009b) worked with a minimum number of 50 presence records for two pine species with a broad geographic distribution range throughout Southeast Asia.

Modelling a species' natural distribution is done under several assumptions, the most important being (1) the species should be in a state of equilibrium with its environment (in other words, the environmental ranges are restricted by competition and predation, and not by dispersion limitations); and (2) the available environmental variables (e.g. climate variables) used in the modelling are determining a-biotic factors in shaping the natural distribution of the species. In practice, one or both of these conditions are often not met. Nonetheless, distribution modelling is still a useful tool for approximating the distribution of a species and, as such, is very relevant to support conservation activities.

Because model outcomes are an approximation of the species' real distribution, it remains a challenge to estimate how representative modelled distributions are. Moreover, the outcomes of distribution modelling can vary depending on the modelling program used, the quality of the presence records and included environmental layers. The outcomes of these models, although potentially useful, should therefore be validated carefully before applying them for *in situ* conservation planning and targeted collecting (Loiselle et al. 2003). There is extensive literature about methods for model validation (e.g. Araújo et al. 2005; Beauvais et al. 2006). DIVA-GIS includes an option to calculate two frequently used indicators of model evaluation – maximum Kappa values and Area Under Curve (AUC) of Receiver Operation Curve (ROC)– from cross-validating modelled distribution maps with a subset of the presence records (Hijmans et al. 2005a). Maxent also provides an option to calculate AUC (Phillips 2009), albeit it is argued that other indicators are more appropriate to measure model performance; for example the separate calculation of omission and commission errors (see Lobo et al. 2008). In chapter three of this work, we present an alternative form to validate the performance of distribution models, i.e. on the basis of formalized expert feedback.

Over the years, a wide variety of ecological distribution models have been described in literature, an exhaustive presentation of which is beyond the scope of this chapter. In the following section, we give a brief overview of the most popular empirical distribution models. The latter are based on observed data and which assume an equilibrium state of the ecosystem, partly based on Peters (2008).

- **Linear Regression models:** regression analysis aims at predicting the pattern in one response variable from the pattern of one or several independent or predictor variables (Kindt and Coe 2005).
- **General linear models (GLMs):** general linear models were developed for situations when certain aspects of the linear regression model are not appropriate. GLMs provide ways of realistically estimating a function of the mean response (the so-called link function) as a linear combination of a given set of predictor variables (Dobson 2002; Nelder and Wedderburn 1972). Popular GLM models are the Poisson GLM with a logarithmic link function (when data are counts) and the binomial GLM with logit link function (for presence-absence data) (Kindt and Coe 2005).
- **General additive model (GAM):** the general additive model extends the GLM by fitting nonparametric smoothing functions to estimate relationships between the response and the predictive variables (Hastie and Tibshirani 1986). The smoothing function generates a curve that can flow more freely between the data than a straight line.
- **Tree-based techniques:** tree-based techniques partition the predictor (environmental) space into parts and then fit a simple model to each part. Classification (categorical response) and regression (continuous response) trees (CART) (Breiman et al. 1984) are a popular technique. Other methods use rule-based classification (Lenihan and Neilson 1993) and maximum likelihood classification (Franklin and Wilson 1991). Random Forests is a related technique that differs from ordinary tree-based techniques in that it generates an ensemble of trees instead of a single best tree (Breiman 2001).
- **Bayesian techniques:** Distribution models based on Bayes' theorem modify an initial (*a priori*) estimate of the probability of encountering a species or vegetation type in a certain landscape by using (1) known preferences (e.g. based on expert knowledge or the literature) of the species or vegetation type for certain environmental characteristics; and (2) information concerning the distribution of these characteristics in the landscape (Tucker et al 1997; Guisan and Zimmerman 2000). However, the quality of the *a priori* information largely determines the model's performance.

As mentioned above, this list of techniques is not exhaustive; many others exist, including artificial neural networks (Lek and Guegan 1999), support vector machines (Cortes and Vapnik 1995), the environmental envelope (Busby 1991) and maximum entropy (Phillips et al. 2006; Elith et al. 2011) models.

In terms of software packages, Maxent, which implements a maximum entropy modelling approach, has performed very well in comparison to others (Elith et al. 2006; Hernandez et al. 2006). It has been used to evaluate the outcomes of species distribution models under different sets of environmental layers (Blach-Overgaard et al. 2010) and to compare the outcomes of species distribution models with the use of different presence record datasets (Feely and Silman 2011). Integrated into DIVA-GIS there are two other distribution modelling programmes: BIOCLIM and DOMAIN (Carpenter et al. 1993; Hijmans et al. 2005a). Although their statistical algorithms are easier to understand than the one used by Maxent, they have not performed that well in comparative studies (Elith et al. 2006; Hernandez et al. 2006). An advantage of Maxent and DOMAIN is that they allow the inclusion of both continuous variables (e.g. climate data), and categorical variables (e.g. layers of vegetation and soil types). BIOCLIM only allows the inclusion of continuous variables. Table 1.6 lists some software packages that are commonly used for distribution modelling.

Table 1.6. Distribution Modelling Packages.

Software	Properties and applications	Source
Maxent	Maximum entropy approach for distribution modelling	www.cs.princeton.edu/~schapire/maxent
BIOMOD	Ensemble forecasting of species distributions, enabling the treatment of a range of methodological uncertainties and the examination of species-environment relationships	http://r-forge.r-project.org/projects/biomod/
OpenModeller	Cross-platform environment where a fundamental niche modelling experiment can be carried out; a number of algorithms are provided as plug-ins, including GARP, Climate Space Model, Bioclimatic Envelopes, Support Vector Machines and others	http://openmodeller.sourceforge.net/
Biomapper	A kit of GIS and statistical tools designed to build distribution models and maps for any kind of animal or plant species centred on the Ecological Niche Factor Analysis (ENFA), which does not require absence data	www2.unil.ch/biomapper/what_is_biomapper.html
DOMAIN	Can operate effectively using only records and a limited number of biophysical attributes	Carpenter et al. 1993

Random Forests	(See text above)	www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm
GARP	The Genetic Algorithm for Rule-set Prediction (GARP) is a distribution modelling method	Stockwell and Peters 1999

Genetic structure and genecological zonation

Spatial patterns of genetic structure are traditionally visualized on spatial data maps by means of vector point data in different colours (e.g. Motamayor et al. 2008) and in pie charts (Trognitz et al. 2011). Pie charts are also used to display similarities and differences in the composition of chloroplast or mitochondrial DNA of different populations (Pautasso 2009). More recently, grid-based diversity analyses with molecular marker characterization data have been used to develop detailed conservation strategies for PGR (Kiambi et al. 2008; van Zonneveld et al. 2012; Thomas et al. 2012) and to understand the origin and dispersal patterns of crops (van Etten and Hijmans 2010).

For most tree species, landraces and CWR, information concerning patterns of intra-specific diversity across their distributions, which would help in prioritizing areas for *in situ* conservation and germplasm collection, is not yet available. In such cases, genecological zonation can provide guidance with respect to the establishment of networks of conservation stands (Graudal et al. 1995). A genecological zone can be defined as ‘an area with sufficiently uniform ecological conditions to assume similar phenotypic or genetic characters within a population’ (Graudal et al. 2004). Under this assumption ecogeographic variation can be used as a proxy of the adaptive genetic variation patterns across a species distribution range (Ramanatha Rao and Hodgkin 2002; Byers 2005). Information about climatic and ecological parameters and topographic barriers can be used to define these zones, which putatively correspond to differences between species populations. The latter are likely to be genetically distinct because gene flow was limited and/or local adaptation to specific environmental conditions.

DIVA-GIS can, for instance be used to map climate zones on the basis of the WorldClim dataset with the use of the clustering option (Hijmans et al. 2005a). Topographic barriers can be visualized with GIS and used to assign species presence records to different populations separated by mountain ranges or water division lines (see table 1.2). Such theoretically constructed zones should ideally be validated by empirical data (ground truth) in order to allow adjustment or refinement of preliminary results. When genetic (molecular or phenotypic) data exist, clustering or ordination techniques can be used to evaluate how much of the genetic structure can be explained by grouping plant individuals in populations according to genecological zones (e.g. Zhang et al. 2006).

Ecogeographic distribution data of specific taxa can provide indicators of their plasticity and adaptation. They can be a useful complement to morphological and molecular diversity studies (or even serve as proxy if morphological or molecular data are not available). In this context, ecogeographic studies support the prioritization of material to be secured in genebanks (Parra-Quijano et al. 2011) and the establishment of core collections for breeding purposes. In chapter 4, we present a case study on ecogeographic analyses of wild potato species endemic to Bolivia.

A programme like Powercore allows the inclusion of ecogeographic variables, such as climate (continuous) and watersheds (nominal), to calculate a subset that is assumed to be representative for a specific taxon in the complete collection (Kim et al. 2007). Ecogeographic studies are also used for carrying out gap analyses. The higher the level of spatial coverage that such a program captures, the greater the amount of genetic variation that is likely to be captured.

In recent years, methodologies and approaches for assessing gaps in genebank collections and prioritizing taxa to be searched in collection missions have been developed. Maxted et al. (2008) provide a geographic gap analysis approach based on a combination of taxonomic, genetic and ecogeographic diversity.

When specific accessions from a genebank collection have shown some interesting traits in evaluation trials (such as drought tolerance, or pest and disease resistance), it can be worthwhile to evaluate genebank accessions collected in the same ecological zone since they will most likely have adapted to a similar environment and might express similar interesting traits. This approach, called Focused Identification of Germplasm Strategy (FIGS) has been used, for example, to pre-select 1320 accessions from a wheat collection of about 16,000, to screen on resistance to powdery mildew. Sixteen percent of these pre-selected accessions eventually showed resistance to the disease (Bhullar et al. 2009).

Monitoring trends in biodiversity

Information about species distribution can be used as an indicator to assess the conservation status of the natural populations of specific plant species. It can be anticipated that species with a narrow and/or fragmented natural distribution are more vulnerable to threats such as changes in land use and climate than species with an extensive and continuous distribution. The International Union for Conservation of Nature (IUCN) has developed several Red List parameters that are based on species distribution, most notably, the extent of occurrence and area of occupancy (IUCN B criterion). In combination with criteria about observed or expected trends in population size, these parameters provide information about the conservation status and threat category of species (IUCN 2010). Distribution-based Red List parameters can be calculated with freely available GIS tools (Willis et al. 2003). In combination with information from species experts on demographic dynamics in and threats to the species plant populations, they can then be used to evaluate the conservation status of wild species, including crop wild relatives (e.g. VMABCC and Bioversity International 2009). In chapter 4, in our ecogeographic study on wild potato species endemic to Bolivia, we also assess their conservation according to IUCN threat categories

It is generally accepted that the modernization of agriculture and changes in land use could have a negative effect on the diversity of crop species and their wild relatives (agricultural biodiversity) conserved on farms and at landscape level. This in turn might lead to genetic erosion at either crop, variety or allele level (van de Wouw et al. 2010a). Nevertheless, exact, well-quantified figures and evidence of genetic erosion as a consequence of agricultural modernization are scarce. Indeed, under certain conditions, crop diversity might even increase when modern varieties are introduced (Bioversity International 2009). Therefore, it is important to (1) establish adequate indicators; and (2) identify areas where agricultural biodiversity can be monitored. Genetic erosion of crops in specific study areas can be measured by comparing current *in situ* diversity with the diversity of genebank material collected from the same area in the past (de Haan et al. 2009a). Current genetic diversity can be compared between different types of land use, such as commercial agriculture versus traditional farming to understand the dynamics in the use of crop diversity (van Heerwaarden et al. 2009). GIS are a useful tool to overlay areas of documented high crop diversity with thematic maps that provide information about accessibility, ethnicity and land use, among other variables, and to better understand which social and economic variables drive the dynamics in the use of crop diversity (e.g. Willemen et al. 2007).

In terms of indicators, those developed within the IUCN criteria may not be appropriate for monitoring dynamics in the use of crop diversity since they are limited to monitoring at species level (rather than intra-specific level). An indicator proposed by the Conference of the Parties to the Convention on Biological Diversity (CBD) (www.cbd.int) is the total number and share in main crop varieties, but this might not adequately reflect relative changes in crop diversity (Eaton et al. 2006).

From a scientific point of view, allelic evenness and richness measured through molecular markers are more appropriate for detecting changes in crop diversity (Eaton et al. 2006). Although molecular diversity studies have become increasingly common and can be applied to monitoring trends in agricultural biodiversity, other, non-molecular-based indicators are also recommended (OECD 2003). These include the share of land devoted to non-intensive production/high biodiversity (with varieties specific to such production systems), percentage of seed of three major crops/varieties originating on-farm and number of traditional (low-production) varieties stored in a genebank (Eaton et al. 2006).

Monitoring of crop diversity on the basis of commercial and traditional varieties might be particularly relevant for specific crops when a sound inventory of registered varieties is in place (Eaton et al. 2006) or when taxonomic keys to distinguish between crop varieties are defined and accepted, such as in the global project of native maize (Proyecto global de maíces nativos, www.biodiversidad.gob.mx/genes/proyectoMaices.html). However, registration of varieties according to morphological characterization may still lead to a substantial degree of misidentification (van de Wouw et al. 2011; Vigouroux et al. 2008). Since the results for characterization with molecular markers are more consistent compared to morphological characterization and quantitative genetics, standardized sets of these markers (Van Damme et al. 2011) are recommended as indicators for monitoring crop genetic diversity (van de Wouw et al. 2010a). For molecular diversity studies, young shoots or other vegetative material from individual plants can be collected in the field and simply stored in bags (such as Ziploc® bags) with silica gel before they are sent to a laboratory for molecular analysis.

The disadvantage of molecular characterization is that in many cases, latent diversity is measured within a sampling unit and not, directly, the diversity of genetic resources (i.e. genetic material of current and future use). Although it can be anticipated that in areas with high neutral diversity, there is also a higher likelihood of finding a high diversity of genetic resources (see next chapter for further details), it is worthwhile to include indicators that directly measure the diversity in traits of interest as well (e.g. morphological descriptors, functional molecular markers, quantitative genetic variation). Similarly, taxonomic identification remains important. Since this is the basis for delimiting the gene pool under study and is essential for identifying target taxa during field collecting, it should be combined with monitoring using molecular marker techniques.

Implementation of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) (www.planttreaty.org) by a growing number of countries calls for increased need in developing an integrated, effective, efficient, global approach to conserving PGR for food and agriculture as part of a rational global system.

Molecular and other types of indicators for analysing agricultural biodiversity (like those described above) are crucial for improving the extent to which variation can be determined in existing *ex situ* collections or under on-farm conditions. They can become a powerful tool for planning new and cost-effective collecting missions (Ramanatha Rao and Hodgkin 2002). In the next chapter of this work, a more detailed overview is presented about the use of molecular characterization data in geospatial analysis to optimize and monitor *in situ* conservation of plant genetic resources.

2. Application of molecular marker technology in spatial analysis to optimize *in situ* conservation of plant genetic resources[i]

Maarten van Zonneveld[1,2], Ian Dawson[3], Evert Thomas[2], Xavier Scheldeman[2], Jacob van Etten[2], Judy Loo[4] and José I. Hormaza[5]

[1]Ghent University, Faculty of Bioscience Engineering, Gent, Belgium
[2]Bioversity International, Regional Office for the Americas, Cali, Colombia
[3]The World Agroforestry Centre, Headquarters, Nairobi, Kenya
[4]Bioversity International, Headquarters, Rome, Italy
[5]Instituto de Hortofruticultura Subtropical y Mediterránea, (IHSM-UMA-CSIC), Estación Experimental La Mayora, Algarrobo-Costa, Málaga, Spain

Abstract

Recent development of new powerful molecular tools that reveal many genome-wide polymorphisms has created novel opportunities for assessing genetic diversity of plant species, especially when these markers can be linked to key adaptive traits and are employed in combination with new geospatial methods of geographic and environmental analysis. New methods to prioritize varieties, populations and geographic areas for *in situ* conservation, and to enable monitoring of genetic diversity over time and space, are now available to support *in situ* germplasm management of annual crop and tree genetic resources. We will discuss concepts and examples of application of molecular marker techniques and geospatial analysis in diversity studies to optimize *in situ* conservation.

2.1 Introduction

There is an increasing need to assess *in situ* diversity status and dynamics of plant genetic resources (PGR) (e.g. in wild populations and on farm) to prioritize and optimize conservation actions and link these effectively with *ex situ* preservation approaches (Palmberge-Lerche 2008; FAO 2010a; 2011). *In situ* PGR are often threatened by modernization and expansion of agriculture, which involves clearance of more land, replacement of landraces by advanced crop varieties, and the practice of new

---

[i] This chapter is used in an adapted form as part of: van Zonneveld M, Dawson I, Thomas E, Scheldeman X, van Etten J, Loo J, Hormaza, JI. Application of molecular markers in spatial analysis to optimize *in situ* conservation of plant genetic resources. In: Tuberosa R, Adler A, Frison E (eds) Advances in genomics of plant genetic resources. Springer. *Accepted*.

management approaches that exclude diversity from the agricultural landscape and lead to genetic erosion (van de Wouw et al. 2010a).

The formulation of effective *in situ* conservation strategies can be optimized by an understanding of spatial patterns of genetic diversity (Petit et al. 1998). Areas of high genetic diversity should be targets for *in situ* conservation as they are considered more likely to contain interesting materials for crop and tree improvement. Measuring genetic diversity is a means for prioritising accessions for further study and use in *ex situ* collections (Frankel et al. 1995a; Tanksley and McCouch 1997; Odong et al. 2012). The approach is now also being used in combination with geospatial methods for presenting results to optimize *in situ* conservation interventions (Samuel 2012; Thomas et al. 2012; van Zonneveld et al. 2012).

Gap analyses that compare the genetic diversity that is present *in situ* with what is maintained *ex situ* provide guidance in devising sampling strategies for the supplementation of *ex situ* collections (Samuel 2012; van Zonneveld et al. 2012). Similarly, comparisons of farm stands with remaining wild plant populations can demonstrate the relative effectiveness of cultivated and natural landscapes for conservation (e.g. Hollingsworth et al. 2005; Miller and Schaal 2005). At the same time, insights in patterns of diversity in the wild and on farmland allow us to better understand the role of evolutionary processes in the development of current species distributions and, where relevant, in their domestication (e.g. Russell et al. 2011). Of course, monitoring activities are also required to measure the effectiveness of *in situ* conservation programs over time, and to account for dynamic processes in the use and management of natural and agricultural landscapes and the transition between the two.

Initiatives that promote the conservation and sustainable use of plant genetic resources draw attention to the need for more assessments of genetic variation in plant species. Molecular characterization is one of the methods prescribed in this context (FAO 2010a, 2011). The recent development of new powerful molecular tools that reveal many genome-wide polymorphisms has created novel opportunities for assessing genetic diversity. This is especially the case when these markers can be linked to key adaptive traits and are employed in combination with new geospatial methods of geographic and environmental analysis (e.g. Escudero et al. 2003; Manel et al. 2003; Holderegger et al. 2010; Chan et al. 2011; Tuberosa et al. 2011). New methods to prioritize varieties, populations and geographic areas for *in situ* conservation, and to enable monitoring of genetic diversity over time and space, are now available and can and should be exploited to improve *in situ* germplasm management.

2.2 Application of molecular marker technology to optimize *in situ* conservation

In situ conservation programs should seek to conserve functional genetic variation that is important to foster future adaptive responses in natural landscapes and to support human needs such as food security and agricultural productivity in managed ones. Often, though, the variation that will be important in the future is not known at the current moment. As a result, some practitioners have therefore taken the view that simply as much variation as possible, whether of known value or not, should be conserved (e.g. van Zonneveld et al. 2012). In this situation, 'neutral' molecular markers that are not linked to any particular trait but presumably providing a representative sample of the 'underlying' diversity in an organism, are appropriate. Although there is little evidence for their association with fitness or adaptive potential (Avise 2010; Ouborg et al. 2010), such markers can provide a good understanding of many of the evolutionary processes involved in the development of contemporary patterns of variation, including historical processes such as contraction and expansion of populations and the development of refugia. They are ideal for understanding the type of species' mating systems, the level of inbreeding and other key features of the biology of species that determine PGR management (Brown and Hodgkin 2008). Such markers also reveal the level of kinship between different crop landraces and the degree of the genetic contribution of different ancestors (Eaton et al. 2006). This has for example been used to prioritize livestock breeds for *in situ* conservation on the basis of their genetic distinctiveness to other breeds (Eding et al. 2002). These methods are now also being applied to crop (maize [*Zea mays* L.]) and tree (cacao [*Theobroma cacao* L.]) genetic resources (Samuel 2012).

Allelic richness at neutral loci is often regarded as an indicator of effective population size (Widmer and Lexer 2001; Leberg 2002). Frankel et al. (1995c) defined the latter as 'the size of an ideal population whose genetic makeup is affected by genetic drift as is the genetic makeup of a populations'. It expresses the rate of historic gene flow and bottleneck events. For this reason is has been used to target wild tree populations for *in situ* conservation (Petit et al. 1998), and to monitor erosion in crop gene pools (van de Wouw et al. 2010a). The number of locally common alleles (alleles that only occur in a limited area of a species distribution but reach relatively high frequencies in these areas) has been identified as a particularly useful measure of richness. Their maintenance at high frequency in particular geographic areas may reflect long processes of selection and adaptation in these specific areas (Frankel et al. 1995a; van de Wouw et al. 2010a). Such areas include centres of crop origin and glacial refugia of wild species which harbour high levels of genetic variation. Identification of areas where geographically restricted alleles occur in high frequency can also be calculated via Allelic Aggregation Index Analysis (AAIA) (Miller 2005). This analysis calculates for a sampled individual on average the nearest geographic distance of its alleles to a similar allele in comparison to the average distance based on the distribution of all individuals (Miller 2005). When only alleles are included in AAIA with a frequency higher than 5 % or more, this function can be an appropriate way to calculate locally common alleles.

In some cases, even with ‘neutral’ markers that are generally not directly related to any function, parameters such as heterozygosity can correspond with population fitness, especially for outcrossing insect-pollinated species (Reed and Frankham 2003; Vranckx et al. 2011). Ideally though, phenotypic evaluation in common-garden experiments should also be applied to understand functional variation in plants. At first sight paradoxically, increases in morphological variation in key features that are selected by humans in the domestication process of annual crops (e.g. *Brassica,* maize, chili peppers [*Capsicum* spp.], potato [*Solanum* spp.]) and trees (e.g. cacao, apple [*Malus domestica* Borkh.]) are often accompanied by decreases in genetic variation in the wider genome. This apparent paradox has fascinated students of domestication for many years. It may be due to bottlenecks induced by human transport of germplasm and/or phenotypic selection events. In both cases sampling only concerns a small part of total genetic variation present in a population. In the case of human selection, the range of variation at traits of interest becomes wider, but elsewhere bottlenecks are introduced (e.g. de Haan et al. 2009a).

Different types of characterization, including different types of markers, thus provide us different types of information and insights. Different marker types may be used simultaneously to target areas for *in situ* conservation because each reveals different features about plant populations. While some may specifically reveal the results of recent gene flows, others may shed light on ancient evolutionary processes that may be related to past climatic fluctuations over tens and hundreds of thousands of years (Newton et al. 1999). Increasingly, markers are being found to be linked to genes that are associated with adaptive traits, which bridge the gap between genetic variation in the genome and form and function. To detect shifts in loci linked to adaptive traits under selection pressure, it is recommended to evaluate these changes against neutral reference loci to (1) distinguish ‘real’ adaptive genetic changes from shifts that are caused by migration and drift; and (2) separate plastic from genetic responses (Hansen et al. 2012). We return to this topic later in this chapter.

The use of molecular tools to target areas for *in situ* conservation rather than just morphological characterization and quantitative genetics has a number of practical advantages. First, it is relatively easy to collect samples needed for molecular analysis in the field and transport them to the laboratory for testing (e.g. it is easier to sample leaves than collect seeds that are only available during a small part of the year and may be recalcitrant or difficult to germinate). Secondly, samples can be analysed in a local laboratory or in another country. For non-living materials, this does not much matter assuming the necessary permissions to exchange genetic material have been obtained. This is particularly relevant when examining species´ diversity patterns across extensive distribution ranges covering many countries. In such cases, all samples can be analysed in one laboratory to assure consistency in materials and analytical methods. Thirdly, markers are neutral to environmental 'noise' that is always present when contrasting the morphological traits of plants that are characterized *in situ* in different locations under different growing conditions. The latter leads to plants that looks different from each other even when they are genetically very similar. The alternative approach is to characterize plants in environment-controlled field trials but these are often expensive whereas a certain amount of environmental noise remains present. Fourthly, modern molecular marker methods are generally repeatable over time and location. This provides opportunities to add data from extra, freshly sampled populations to existing data sets. The latter is especially practical to monitor the dynamics of diversity in populations over time, for example, in assessing genetic erosion. The molecular diversity of a 'historic' collection from a specific area can be compared with a new one, such as de Haan et al. (2009a) did to assess allelic loss over time in local potato varieties grown in Peruvian Andean villages. In this particular case, no loss of molecular diversity was observed over a 25-year period, suggesting on-farm conservation with farmers was effective. When improved varieties cross with local landraces and are taken up in informal seed systems they may reduce *in situ* diversity as shown for maize in southern Mexico, even though these varieties also introduce molecular and phenotypic variation in these systems, which are different from the genetic diversity found in landraces (van Heerwaarden et al. 2009).

Despite the reductions of *in situ* crop genetic diversity due to replacement by improved varieties and crosses with local varieties, levels of newly introduced variation may increase as well. Meta-analysis of molecular diversity studies of eight food crops suggest that in the last decades breeders have increased the use of crop diversity in the development of improved varieties (van de Wouw et al. 2010b).

As mentioned in the introduction, on-farm conservation can also complement *in situ* conservation of wild populations that are increasingly under pressure due to deforestation and other threats. This is relevant for many locally socio-economic important tree species that are in incipient phases of domestication. However, further research should be carried out to better understand under which ecological and socio-economic circumstances on-farm conservation could be an effective approach for sustainable and long-term tree genetic resources management (Dawson et al. 2013). In the case of the Amazonian ice-cream bean tree (*Inga edulis* Mart.), for example, nuclear SSR diversity is lower in farms (e.g. in backyards and living fens) compared to that of wild populations, although allelic variation remains relatively high in the former whereas agricultural landscapes are still important sites for conservation (Hollingsworth et al. 2005). In another example, cultivated populations of the Mesoamerican fruit tree species jocote (*Spondias purpurea* L.) contain unique chloroplast alleles that are not found in wild populations (Miller and Schaal 2005). This shows how on-farm conservation complements conservation of wild populations and suggests that farmers can preserve genetic variation that otherwise is lost due to past and current decline of natural populations (Miller and Schaal 2005).

2.3 Geospatial analysis techniques for mapping molecular genetic diversity

Just as molecular marker methods have advanced greatly over the last decade, so too have approaches for geospatial analysis (Guarino et al. 2002; Miller 2005; Jombart 2008; van Etten and Hijmans 2010; Chan et al. 2011; van Zonneveld et al. 2011). Advances in computational applications for geographic information systems (GIS) and spatial analyses are still however underutilised in genetic diversity studies. This may be because many scientists are unaware of the newer methods available. Training materials have been developed to bridge this gap (Scheldeman and van Zonneveld 2010). A great advantage of GIS-based approaches is the clear graphic presentation of results through maps. The latter facilitates the interpretation of findings and hence their incorporation into conservation strategies (Jarvis et al. 2010). Geospatial analysis of genetic diversity has been undertaken for a wide range of tree species because the maintenance of genetic resources of most of these species depends largely on *in situ* conservation. For the Norway spruce (*Picea abies* [L.] Karst.) in Austria, for example, a geographic grid-based gap analysis has been carried out to identify new conservation units that complement the coverage of mitochondrial and nuclear molecular marker variation, and adaptive genetic diversity by the current species´ conservation unit network (Schueler et al. 2012).

One effective method to describe genetic diversity in geographic space is to use circular neighbourhood-type analyses. This approach is especially effective when working with individually geographically referenced accessions rather than with populations (van Zonneveld et al. 2012). The circular neighbourhood allows calculating confident genetic parameters per grid cell by grouping georeferenced individuals with each other within a user-defined radius of geographic distance around each grid cell (Scheldeman and van Zonneveld 2010). It also makes analyses less sensitive to grid origin definition and enables the inclusion of isolated trees in the calculation of the genetic parameters, i.e. together with their closest neighbouring trees.

As a second step after circular neighbourhood, uneven sampling densities among grid cells can be corrected by establishing a level of rarefaction (minimum sample size per grid cell to include in analysis) or by carrying out re-sampling without replacement (see Leberg 2002). The final results of the corrected diversity analysis after circular neighbourhood provide then detailed and representative estimates of geographic patterns of diversity for a particular area. The scaling can be related appropriately to the dimensions of particular countries or regions within countries so that results can be incorporated into national and regional conservation plans. Such an approach has been used to identify genetic diversity hotspots for the *in situ* conservation of a number of important perennial tree crops, including the fruit tree cherimoya (*Annona cherimola* Mill.) in the Andes (van Zonneveld et al. 2012), cacao in its Latin-American centres of origin and domestication (Thomas et al. 2012), and the bush mango (*Irvingia gabonensis* (Aubry-Lecomte ex O'Rorke) Baill. and *I. wombolu* Vermoesen) in Central Africa (Lowe et al. 2000). In chapter 6 of this work, the application of circular neighbourhood and rarefaction is illustrated to map microsatellite marker diversity using the case of cherimoya.

These are examples of geospatial analysis to prioritize areas for the conservation of genetic resources of just a few economically important trees. However, thousands of tree species have local livelihood value whereas others play important ecological roles in local ecosystems, and many of these are threatened. As the costs to carry out analyses with different molecular marker types increasingly are reducing, it will become more and more feasible to perform such studies for more tree species.

One approach to extrapolate patterns observed from these analyses and prioritize areas as many tree genetic resources as possible is to identify Pleistocene refugia and converging post-glacial migration routes. These areas harbour high inter-specific and intra-specific diversity (Petit et al. 2003). Georeferenced observation points of such species from herbaria and genebanks can be used to predict Pleistocene species distributions on the basis of past climate data (Waltari et al. 2007). Such data is freely available from the PMIP2 website (http://www.pmip2.cnrs-gif.fr) although it still needs to be down-scaled. Georeferenced plant data and climate models are increasingly available through online platforms such as the Global Biodiversity Information Facility (www.gbif.org) and WorldClim (www.worldclim.org), respectively. These data, where available and when they are of reasonable quality, can be fed into Environmental Envelope Modelling (EEM) to predict past species distributions and reconstruct potential Pleistocene refugia (Waltari et al. 2007; Thomas et al. 2012; Vinceti et al. 2013). Available neutral molecular marker data, especially those that detect chloroplast DNA variation, can help to validate or refute these potential refugia from which species dispersed after glacial periods (Newton et al. 1999; Petit et al. 2003). They also help to detect converging migration routes that have high levels of diversity due to hybridization of populations with distinct ancestries (Petit et al. 2003). A major limitation, however, is that often different sampling methods, markers and marker types are used in separated studies of the same species in different parts of its distribution range. This complicates comparability and clear identification of distribution-wide diversity patterns, for example for the new world tropical palm species *Bactris gasipaes* Kunth (Clement et al. 2010; Graefe et al. 2013). For most important food crop species already standardized molecular tool kits have been proposed to improve comparability in genetic characterization (Van Damme et al. 2011). However for most other economic plant species, molecular standards still need to be developed.

It may also be possible to make species-specific inferences based on general principles that apply across crops, without having full genetic data. Crops were domesticated only about 13,000 years ago or more recently and the current distribution of their diversity is marked by the subsequent process of dispersal. High levels of inter-specific and intra-specific crop diversity can expected to be found in and around centres of crop domestication such as the Andes, Mesoamerica and the Amazon in the Americas. Just as high tree genetic diversity is expected in post-Pleistocene converging migration routes, high crop diversity can also be expected in converging dispersal routes.

An example is cultivated chili pepper genetic diversity in Peru. The diversity of cultivated *Capsicum* encountered in Peru is probably the highest in the world (Meckelmann et al. 2013). It is an important area of diversification and varieties from the five cultivated species are cultivated since early pre-Columbian times (Perry 2012). However, Peru is probably not the centre of crop origin of any or most of the five cultivated species. Consequently, they had to be transported to Peru from their putative centres of domestication in respectively Mexico (*C. annuum* L.), Bolivia (*C. baccatum* L. and *C. pubescens* Ruiz and Pav.) and the Amazon including the Peruvian part (*C. chinense* L.). The crop origin of the fifth domesticated species, *C. frutescens* L., is less clear. It occurs in tropical Mesoamerica and the Amazon (Eshbaugh 2012).

Studies in human genetics show that relatively simple models of diffusion can be used to predict global genetic diversity of any living being (Ramachandran et al. 2005). Diffusion models have also been used to model the spread of agriculture or particular crops (Pinhasi et al. 2005). van Etten and Hijmans (2010) showed that for crops, spatial diffusion models and genetic diversity models can be linked. Such models could eventually be used to predict levels of diversity and complementarities between locations, including un-sampled locations, based on inferences on the underlying historical processes that created the spatial diversity pattern to start with.

Besides studies across crop distribution ranges to prioritize areas for *in situ* conservation, spatial studies with a more local scope in targeted areas can be important to decide in more detail about the most appropriate on-farm PGR management strategies in traditional rural communities. Such studies can, for example, help to better understand how farmers manage and conserve crop diversity within the landscape and through time (Worthington et al. 2012), and to identify at which geographic and social level *in situ* conservation should be implemented and crop diversity monitored (Barry et al. 2007). Estimates of distribution and levels of crop diversity in rural communities also help determining the need to introduce new varieties into local seed systems and improve seed distribution systems accordingly (Jarvis et al. 2011).

PGR management in traditional rural communities differs per crop species and social context. For example, microsatellite markers showed that farmers in southern Mexico maintain bean species diversity (*Phaseolus coccineus* L.; *P. dumosus* Macfad.; *P. vulgaris* L.) and *P. vulgaris* landraces clearly separated in different fields along a topographical climate gradient (Worthington et al. 2012). A microsatellite marker analysis of the genetic structure of rice (*Oryza sativa* L., *O. glaberrima* Steud.) in SDR Guinea revealed genetic differences between lowland coastal and upland savannah agro-ecosystems but no differentiation between villages or farms within each of the contrasting agro-ecosystems (Barry et al. 2007). Although within in each variety, high genetic diversity was found, most of this diversity can be conserved within just a few farms of a village (Barry et al. 2007). The high diversity within farms and low genetic structure between farms can be explained by active human seed exchange and high varietal turnover (Barry et al. 2007). Likewise, in Peruvian Andean potato growing areas, most variation in microsatellite diversity in potato (*Solanum* spp.) is observed to be kept within farmer families (de Haan et al. 2009a). No much genetic differentiation was observed between two geographically distant areas (de Haan et al. 2009a). However, whereas in some communities many of the farmer families still conserve high varietal diversity, in other villages, only a few families conserve large diversity (de Haan et al. 2009a). Conservation of high potato varietal diversity can be explained by preferences of specific cultivars for own consumption and risk spreading through varietal diversification (de Haan et al. 2009a,b)

As some of the above examples already illustrate, GIS can be used to overlay whole series of different information types onto genetic data to make more informed resource management decisions. The latter include drivers of genetic erosion of natural species populations, such as threats to ecosystems and their relative vulnerability (e.g. Jarvis et al. 2010; Hirota et al. 2011). But there may also be factors that are responsible for an increase in genetic diversity of natural and cultivated populations such as increased landscape connectivity and seed exchange.

Threats to ecosystems need to be interpreted carefully when applied to specific species, since individual taxa, and populations within them, will respond differently to them. Nevertheless, areas of important genetic diversity under threat can be identified for urgent conservation, such as important natural populations with high allelic richness that are located in areas with agricultural encroachment and/or important contemporary populations in locations where future climate will likely not support regeneration and survival. At the same time, low-threatened populations with high genetic diversity can be identified. In these areas, relatively low investments would already result in adequate *in situ* conservation.

Recent studies have begun to explore how to include spatially defined threat information to prioritize varieties for *in situ* conservation. This allows calculating optimal solutions for conservation interventions considering the costs to conserve each variety and the unique genetic diversity that would be maintained by conserving these varieties (Samuel 2012). Variables that could serve as a proxy to the chance of variety replacement include like geographic distances to populated places, and climate change impacts. Genetic diversity information can consist of cacao and maize varieties as delineated by molecular markers (Samuel 2012).

EEM of species distributions within current and projected future climates and assessing the changes in distribution ranges between time intervals can be used to identify hotspots of genetic diversity that are particularly vulnerable to change. This has for example been done for cacao (Thomas et al. 2012). The comparison of current and future modelled distribution areas revealed several areas of low climate change threats within the Amazonian area of high cacao genetic diversity that should be targeted for *in situ* conservation (Thomas et al. 2012). Tree species are particularly interesting for these studies because of their longevity. That makes them good candidates for studying climate change impacts on landscapes, and thus for delineating the consequences of anthropogenic climate change (Petit et al. 2008). Generally, the available molecular data in combination with other data sets such as pollen cores would suggest that natural dispersal will not be able to keep up with climate change in many parts of the world. Therefore whole forest ecosystems that are crucial for *in situ* conservation of trees and associated flora (including the wild populations of some crucial crops and their relatives) and fauna may be threatened (Malhi et al. 2009).

For annual crops, a good example of application of current and past climate EEM is the wild subspecies of barley (*Hordeum vulgare* subsp. *spontaneum*) in the Fertile Crescent and Central Asia (Russell et al. *in prep*). In this case, contemporary patterns of molecular marker diversity expressed using the circular neighbourhood method corresponded with EEM for the Last Glacial Maximum. Both these analyses indicated that the eastern Mediterranean was a Pleistocene refugia for the wild subspecies of barley. Barley here has higher genetic diversity in the eastern Mediterranean than elsewhere in its distribution. This area should therefore be a focus of conservation activities for barley genetic resources.

Most interestingly in this case, geographic point location data of barley accessions were used to identify the environments in which the taxon grows in its natural range, by extract values from the 19 bioclimatic variables of WorldClim (www.worldclim.org). The advantage of barley compared to many lesser-studied plants is that the chromosome positions of many molecular markers are known. This kind of landscape or environmental genomics approach based on point- referenced climate data and association genetic techniques has the potential to be very useful in crop breeding and in monitoring responses to environmental change. It is likely to be increasingly adopted more and more in the near future. In the case of the wild barley, for example, by comparing associations between particular chromosome-mapped markers and bioclimatic variable data across sample locations, it was possible to identify regions of the barley genome that are candidates for being part of the responsible genes for adaptive responses to the environment. This type of analysis is especially relevant for plant species for which comparatively little data on the inheritance of important traits are available (Neale and Kremer 2011). Such methods may make data taken directly from wild stands (not only from collected material in field trials) more important in the future for breeding purposes.
The study of genetic and plastic responses of plant populations to local change is especially relevant because migration to more suitable locations may be restricted due to fragmentation and the rapid pace of climate change (Hoffmann and Sgró 2011). At the same time, distribution range shifts may cause reduced fitness of populations due to

founder effects, i.e. genetic reduction in new populations that have been established by a few individuals (Cobben et al. 2011). Landscape genomics and association mapping of DNA variation to geographic environmental patterns hold much promise to monitor adaptive genetic responses over time under progressive climate change (Hansen et al. 2012). Molecular data modelled in geographic space can help determine potential migration rates. Adaptation to current locations can be monitored through allele shifts at important genes (as described in the barley example above). The latter approach is becoming increasingly feasible as chromosome-mapped markers are linked to adaptive traits.

Conservation genomics, i.e. combining conservation genetics principles with functional genomics approaches is, in Ouborg's (2010) opinion both necessary and feasible, to understand the effects of loss of genetic diversity to fitness and adaptive potential. Avise (2010) noted that the 'genomics revolution' allows to examine sequence variation in unprecedented numbers of loci for unprecedented numbers of individuals in and among populations. Although most genomic advances are currently associated with well-studied species, rapid developments will allow for genome-wide mapping in virtually any plant species in the near future (Ingvarsson and Street 2011).

The main challenge for mapping diversity of alleles having adaptive significance may be that many quantitative traits are influenced by many loci. If drought tolerance is influenced by more than 200 loci, for example, what is the value of choosing a handful of these, more or less at random? Also, in studies to date, only a small portion of trait variability has been explained by the tens of loci characterised. This is the so-called 'missing heritability' phenomenon. Unlike neutral alleles, alleles that are under selection pressures which are incompletely understood, have the potential to provide richer information but also to mislead if sample size is insufficient.

3. Application of consensus theory to formalize expert evaluations of plant distribution models for better species´ conservation[i]

Maarten van Zonneveld[1,2], Nora Castañeda[2,3], Xavier Scheldeman[2], Jacob van Etten[2] and Patrick Van Damme[1,4,5]

[1]Ghent University, Faculty of Bioscience Engineering, Coupure links 653, 9000 Gent, Belgium
[2]Bioversity International, Regional Office for the Americas, PO Box 6713, Cali, Colombia
[3]Centro Internacional de Agricultura Tropical (CIAT), PO Box 6713, Cali, Colombia
[4]World Agroforestry Centre (ICRAF), Global Research Programme 1, PO Box 30677, Nairobi, 00100, Kenya
[5]Institute of Tropics and Subtropics, Czech University of Life Sciences. Prague, Kamycka 129, Prague 6, Suchdol, 165 21, Czech Republic

Abstract

Application of Environmental Envelope Modelling (EEM) for conservation planning requires careful validation. Opinions of experts who have worked in the field on conservation, seed collection and ecology of the specific species of interest can be a valuable and independent information source to validate EEM because of their first-hand experience with species occurrence and absence. However, their use in model validation has always remained limited due to the subjectivity of their feedback. In this study, we present a method on the basis of cultural consensus theory to formalize expert model evaluations. We developed for five tree species, distribution models with nine different variable combinations in Maxent EEM software. Species specialists validated the distribution maps generated through an online Google Earth interface with scores ranging from *Invalid* to *Excellent*. Experts were also asked about commission and omission errors of the distribution models they evaluated. We weighed expert scores according to consensus theory. These values were used to get a final average expert score for each of the distribution models produced. Consensus-weighed expert scores were compared with un-weighed scores and correlated to four conventional model performance parameters after cross-validation with test data: Area Under Curve (AUC), maximum Kappa,

---

[i] *Submitted*

commission error and omission error. Median consensus-weighed expert appreciation was close to *Fair*. In general, experts who reached more consensus with peers were more positive about the distribution model outcomes compared to those who had more opposite judgements set against others. Both consensus-weighed and un-weighed scores were significantly correlated to corresponding AUC, maximum Kappa and commission error values. We found no correlation between expert scores and omission errors. More than half of the experts indicated that the distribution model they considered best, included areas where the species is known to be absent (commission) while a smaller but still substantial proportion (31 %) also indicated areas of species presence that were omitted by the model. Methods to formalize expert knowledge allow a wider use of this information in model validation and improvement, and complement conventional validation methods of presence-only modelling. Online GIS and survey applications can facilitate expert consultation.

3.1 Introduction

A good understanding of the actual distribution of any plant species is one of the key parameters allowing evaluation of its conservation status and the formulation of effective conservation strategies. However, for most plant species, only a limited amount of data on their distribution is available (Nic Lughadha et al. 2005; Newton and Oldfield 2008). This is particularly true for regions that harbour high levels of plant diversity, including tropical and subtropical zones in Africa, Asia, Latin America and the Caribbean (Nic Lughadha et al. 2005).

Environmental Envelope modelling (EEM) can be used to develop predictive models that make inferences about species´ geographic distributions (Araújo and Peterson 2012). EEM is therefore considered a useful tool to overcome the lack of complete distribution data (Guarino et al. 2002). This kind of modelling technique defines a species' ecological niche to predict areas of potential species occurrence. This is done on the basis of environmental data obtained for occurrence sites where a species has been observed and from sites where it is absent. Because absence points are difficult to obtain, often randomly generated background points are used as an alternative to discriminate less suitable environments from more suitable environments in areas where the species has been observed (Pearce and Boyce 2006). Presence points can be derived from georeferenced herbarium specimens, genebank accessions and/or vegetation/plant species inventories. The latter are made increasingly available online by herbaria and genebanks through portals such as the Global Biodiversity Information Facility (GBIF) (www.gbif.org). One of the advantages of EEM is that no prior knowledge on the ecophysiology or reproductive biology of plant species is needed to develop a model (Guisan and Zimmerman 2000). This allows a systematic approach for predicting distributions for conservation planning for large species numbers and the assessment of their conservation status.

EEM has several advantages. However, application of this tool in conservation planning should be critically evaluated. To it is, the algorithm chosen to model species distribution from actual observation data influences the outcomes. This may lead to modelled distributions that deviate significantly from reality (Loiselle et al. 2003). An additional challenge comes from the fact that the modelled distribution ranges are influenced by the environmental variables included and/or omitted in the model. An adequate selection of determinant variables for any species´ distribution can thus improve the model significantly (Austin 2007).

The results of EEM presence-only modelling have been extensively cross-validated with test data consisting of presence and pseudo-absence points using statistical parameters such as maximum Kappa and/or Area Under Curve (AUC) (e.g. Loiselle et al. 2003; Elith et al. 2006; Hernandez et al. 2006). Nevertheless, because of the lack of confirmed points of species absence, it remains difficult to provide a good estimate of the commission error, i.e. the extent to which models predict occurrence in areas where the species is actually absent (Anderson et al. 2003; Rupprecht et al. 2011). In addition, observer bias can result in dependence between presence points used to develop a distribution model (training data) and the presence points to will be used to validate the model (test data) (Dorman et al. 2007). This may lead to high rates of model performance whereas a model may actually omit many not-yet-sampled areas of species occurrence (Hijmans 2012).

Opinions of experts, like foresters, ecologists, botanists and park managers are another key information source that can be used to validate and fine-tune the outcomes of EEM because of their experience with specific species in the field (Thuiller 2003; Beauvais et al. 2006). They are probably also a more independent source for model evaluation than cross-validation with test data. They can provide valuable information about the extent to which models predict species absence in areas where species do not occur naturally. They may also be a good source to validate model performance in under-sampled areas. Park managers may validate, for example, species occurrence according to distribution models in protected areas that have not been sampled due to administrative constraints or because collectors had prioritized more threatened areas for sampling. Expert feedback also provides insight on how relevant potential users consider distribution modelling to be for their field activities on *in situ* conservation, seed collection and inventories of specific species, to name just a few potential uses. This fits in a wider discussion about the applicability of species distribution mapping and EEM for conservation and sustainable use of biodiversity (Knight et al. 2008; Araújo and Peterson 2012).

Increased computer capacity and internet availability during the last decade have allowed the development and widespread application of many new, powerful distribution modelling tools to predict species distribution (e.g. Elith et al. 2006; Thuiller et al. 2008). At the same time, this has allowed the development of online tools, such as ArcGIS Server, Google Earth and GeoWiki, which make it possible to remotely consult specialists including botanists, ecologists or park managers located in different parts of a country and in the world, and consider their opinion in distribution model validation and improvement. However, expert-based judgements are often not applied or reported in evaluating EEM because they are considered to be subjective. Indicators on the basis of presence points and pseudo-absence points, like AUC or maximum Kappa, are, despite their limitations, preferred in EEM studies because of their formal nature which allows repeatability and comparability between different studies.

It is possible, however, to analyse expert-based opinions in a more objective way. Romney et al. (1986) developed an approach to formalize informant knowledge on the basis of cultural consensus theory. The consensus model estimates the probability that an informant provides correct answers dependent on the concordance of her/his answers with overall group consensus. It has been applied in social and ethnobotanical sciences to weigh informant responses (e.g. Weller and Mann 1997; van Etten 2006; Benz et al. 2007).

In this study, we present an approach on how to formalize expert evaluation applying consensus theory to select and examine the relevance of distribution models for species conservation assessment and planning. The rate of expert agreement can be used as a degree of confidence of model evaluation and selection of the best distribution model. Secondly, the method allows to identify for each expert how trustworthy his/her answers are on the basis of the consensus he/she reaches with his/her peers. On the basis of these values opinions of different experts from an informant group can be weighed in final model evaluation.

In distribution modelling, expert knowledge has been used to identify critical environmental variables and species environmental ranges in the case of small sample sizes (Barry and Elith 2006) or to identify areas for crop suitability (FAO 2007). It is also being incorporated in the development of distribution models (Bierman et al. 2010). However, we found only a few references that reported the use of experts for model evaluation (Anderson et al. 2003; Ramírez-Villegas et al. 2010). To our knowledge, this is the first time an approach is presented to formalize expert knowledge for EEM outcome validation.

For five socio-economically important tree species native to Latin America and the Caribbean, we present distribution models generated in Maxent with nine different climate variable combinations. Species specialists evaluated model outcomes through an online survey in Google docs with a dynamic Google Earth interface. We compare expert judgements with and without applying consensus theory with four commonly used validation measures on the basis of cross-validated presence and pseudo-absence test data; maximum Kappa, Area Under Curve (AUC), and commission and omission error.

We further examine the patterns of variable selection and model appreciation by experts with and without applying consensus theory.

3.2 Methods

Species

The five tree species we tested here are *Annona cherimola* Mill. (cherimoya), *Bactris gasipaes* Kunth (peach palm), *Bertholletia excelsa* Bonpl. (Brazil nut), *Cedrela odorata* L. (Spanish cedar) and *Nothofagus alpina* (Poepp. & Endl.) Oerst. (raulí). These species were prioritized by LAFORGEN (www.laforgen.org), the Latin American Forest Genetic Resources Network of scientists and practitioners, and have been selected in a project named MAPFORGEN (www.mapforgen.org). This project aims at evaluating the conservation status of 100 socio-economically important woody species native to Latin America and the Caribbean. As part of this analysis, the species distribution ranges are modelled. The five selected species occur in different ecological and geographical zones in Latin America and the Caribbean, and their distribution has been studied relatively well compared to other MAPFORGEN species.

Environmental Envelope Modelling (EEM)

We applied a presence-only EEM approach using the Maxent program (Phillips et al. 2006). This is a widely used distribution modelling tool of which the algorithm is reported as predicting species distribution well, in comparison to other modelling software (Elith et al. 2006; Hernandez et al. 2006). It is already used by several environmental agencies (Elith et al. 2011).

We obtained presence points coming from herbaria and genebanks for the five selected species through GBIF. This dataset was complemented with presence points provided by several members of LAFORGEN, (Corporación para el Desarrollo de los Recursos Naturales [CEDERENA], Ecuador; World Agroforestry Centre [ICRAF] Peru; Instituto Forestal [INFOR] Chile; and Instituto Nacional de Tecnología Agropecuaria [INTA] Argentina). We only considered points within the native distribution ranges defined according to the Germplasm Resources Information Network (GRIN) of the United States Department of Agriculture, Agricultural Research Service, National Genetic Resources Program (USDA, ARS, NGRP) (http://www.ars-grin.gov/). The timber species *C. odorata*, and *N. alpina* occur in general only in natural populations. The distribution of the non-timber species *B. excelsa* is hypothesized to be shaped by human-environment interactions in the Amazonian forest (Sheppard Jr and Ramirez 2011). The fruit species *A. cherimola* and multi-use palm species *B. gasipaes* are in phases of incipient or semi-domestication (National Research Council 1989b; Clement et al. 2010). Their species records include records from natural populations, and *circa situm* observations of trees maintained in backyards, home gardens and smallholder farms.

We checked the observation passport data for inconsistencies between the recorded coordinates and the reported highest-level administrative unit in a country (e.g. departments or states), after Scheldeman and van Zonneveld (2010). Inconsistent points were removed. In addition, we used a Mahalanobis distance analysis to identify points in atypical climates ($0.025 < p > 0.975$) as they are probably errors (Chapman 2005). Distances between points were calculated with values of 19 bioclimatic variables as defined by Busby (1991) representing different interannual bioclimatic conditions important for a plant's natural establishment and survival. Climate data were derived, for each species presence point, from the 30-seconds resolution Worldclim dataset (Hijmans et al. 2005b).

Each of the nine models that we developed in Maxent, used as input a different environmental variable combination from the 19 bioclimatic variables, one soil-type classification map and a categorical ecological zones map (Appendix 3.1). We also selected a core set of four bioclimatic variables that represent different facets of intra-annual climate conditions. This set of variables consisted of annual mean temperature (°C), annual precipitation (mm/y), temperature seasonality (standard deviation of monthly temperature x 100) and precipitation seasonality (variation coefficient of monthly precipitation). The map of soil units was derived from the SOTERLAC database (Batjes 2005) and followed FAO´s classification of soil units (FAO 1988). The map of ecological zones was derived from FAO's terrestrial ecological zone classification (FRA 2001).

We used Maxent default settings when modelling species distributions and applied the 10 percentile training presence threshold to restrict potential distribution areas. This latter is one of the threshold values provided by Maxent and limits the modelled areas of occurrence to a distribution range in which 90 % of the presence points are located inside the modelled area while 10 % of the presence points are outside the modelled areas of occurrence.

Background points were taken from the whole study area that comprises Latin America and the Caribbean (maximum longitude in decimal degrees = -32.375, minimum longitude = -121.125, maximum latitude = 34.5833, minimum latitude = -55.9583). From the modelled areas, we excluded intensive agricultural areas, bare lands and urban areas as delineated by the Global Land Cover 2000 Project (Fritz et al. 2003). We anticipate that our tree species do not occur in these land use types because these areas have low forest cover and no natural vegetation.

Online expert evaluation survey

For each species, we developed an online survey in Spanish (see Appendix 3.2). Hyper Text Markup Language (HTML) code and Cascading Style Sheets (CSS) were used to develop a web page to present a questionnaire for each of the five species. Within the web page of each species respectively, the nine modelled distribution maps were presented in Keyhole Markup Language (KML) format in an Application Programming Interface (API) of Google Earth. An embeddable form hosted in Google Docs was included in the web pages to store the evaluation scores provided by experts. For each species, we sent an invitation with a link to the online survey to: 1) LAFORGEN members who had indicated research interest in conservation and use of the respective species (many of them are actively involved in such research); and 2) researchers who we found to have studied these species, following a literature review of genetic and ecological studies for each respective species. In total, 99 persons were invited to participate. The survey took place from 10 August 2009 to 29 September 2009.

In the Google Earth interface, each respective species expert could select and view the modelled distribution derived from each of the nine variable combinations to evaluate them visually. Experts were asked to concentrate on the areas they knew best. We asked them to indicate their geographic area of expertise (e.g. country and/or departments or provinces). Distribution maps were presented on a scale from low (yellow) to high probability (red) of species occurrence. Experts did not receive information about the environmental datasets that had been used to generate each model. Specialists could zoom to the geographical distribution area of their expertise (we recommended a minimum eye height of about 25 miles (~ 40 km), whereas they could choose one of five scores to rate the modelled distributions: 1 (invalid), 2 (bad), 3 (fair), 4 (good) and 5 (excellent).

Commission and omission errors according to experts

Distribution models used in conservation planning should ideally have a low commission error to minimize the costs for implementing conservation measures to protect species (Araújo and Peterson 2012). Over-predictions result in high rate of commission errors. They can occur because migration limitations to movement are not taken in account in the EEM. They include past and current barriers that can substantially restrict real plant species´ distributions compared to their potential distributions (Svenning and Skov 2007). On the other hand, for the discovery of new populations it is important that models have a low omission error (Araújo and Peterson 2012). Accessing new populations is important for germplasm collecting and maximizing *in situ* plant genetic resources conservation. Omission errors may occur because of sampling bias resulting from over-sampling in areas which are easy accessible, such as areas close to roads. Sampling is much more difficult in more remote areas with potentially new populations, which remain under-represented and may consequently be under-predicted in EEM (Hijmans 2012).

Therefore, we also asked each expert if the model that he/she had selected as producing a distribution most similar to the species distribution in their area of expertise contained commission and/or omission errors. We further asked the reasons for commission error; whether model prediction in areas of species absence was due to human-mediated species extinction and/or because these areas were outside the natural distribution range.

Application of consensus theory to formalize expert evaluation

The consensus model assumes that each informant has a probability to provide the correct answers which are not known to researchers prior to questioning. In our case, we didn't know the real distribution areas of our study species. Nor, did we know how the different distribution models are related to the real species distributions. The model further assumes that respondents come from the same cultural group. Romney et al. (1986) present a hypothetical example of a cultural group consisting of tennis players that use the same jargon vs. a cultural group of non-tennis players who are less consistent in their answers on the rules of game because they have not that much knowledge about this sport. In our case, we tapped into a scientific community of peers. We assume that this community consists of one cultural group, although our experts come from different biological disciplines and were maybe trained with other conceptual backgrounds. A third postulation is that informants' answers are independent from each other (Romney et al. 1986).

The consensus model estimates the accuracy of an informant´s response on the basis of the latter concordance with overall group consensus on this answer in his or her cultural group. The rates of accuracy or competence (between 0 and 1) can then be used to weigh each informant's response in the final analyses. Indeed, the results from several case studies support consensus theory confirming that informants whose answers are closer to consensus also have more correct answers compared to persons whose answers are more divergent from consensus (see Romney et al. 1986).

In this study, we used the rate of agreement between species experts as a way to validate accuracy of the overall expert model evaluation and selection for a specific species. Secondly, we used the expert competence rates to weigh average expert scores per species model and variable combination. In the remaining text of this chapter, we will refer to these scores as consensus-weighed expert scores. Similarly, un-weighed expert scores were calculated, but without taking in account competence values.

We examined how consensus-weighing influences (1) best model selection according to experts; (2) quality of the distribution models in general according to experts; (3) expert score correlation with Maximum Kappa and AUC; and (4) commission and omission errors. The steps involved to calculate competence values are written with the basic functions included in R (R Development Core Team 2010).

The first step in consensus model calculation is the development of a matrix with the proportions of agreement in answers between paired experts. Originally, Romney et al. (1986) developed this matrix on the basis of the rates of matches between 0 and 1 in answers on true/false or multiple choice questions (Romney et al. 1986). Later this has been extended to covariance matrices (Weller and Mann 1995). In our case, each species expert provided a rank score from 1 to 5 for nine different models. Instead of rate of matches or covariance, we then calculated the proportion of agreement between respondents with Spearman correlation coefficients. The main difference between correlation coefficients and rates of matches is that correlation coefficients can also be negative when two experts systematically disagree. Consequently, these coefficients range from -1 to 1.

The second step is correction of matches for guessing (Romney et al. 1986). In our case, the chances that two respondents return the same series of scores by simply guessing are practically zero. However, to avoid singular computations in further analysis of the correlation matrix, we subtracted 0.0001 from the pairwise correlation coefficients.

We then carried out a maximum-likelihood factor analysis on the correlation coefficient matrices. This was only done with one factor, as indicated by Romney et al. (1986). The amount of variance explained in this first factor reflects the rate of consensus between experts (Weller and Mann 1995). We used this as an indicator of the rate of expert agreement on model performance and best model selection.

The results from the maximum-likelihood factor analysis were also used to obtain for each expert a so-called value of competency. Expert scores can only be weighed with zeros or positive competence rates ($0 \leq D \leq 1$). However, an expert could receive negative competence values when (s)he rated consistently opposite to consensus scores. In these cases, values were converted to zero, i.e. the lowest competence value that can be contributed to weigh expert scores.

Experts that gave equal scores to all nine models were excluded from the calculation of consensus-weighed and un-weighed expert scores because this prevented us from knowing which model these experts considered to be the best.

Selection and relevance of variable combinations

We carried out a non-parametric ANOVA test (Friedman) to test if one or more of the nine distribution models were consistently more appreciated by the experts of the five different species compared to the other models. We also examined if there were differences in variable combination appreciation between consensus-weighed and un-weighed expert scores.

Correlation of expert-based judgement with conventional model performance parameters

We compared consensus-weighed and un-weighed expert scores with the corresponding values of four commonly used parameters in EEM outcome validation: AUC, maximum Kappa, and commission and omission error values from cross-validation using Pearson´s correlation coefficient. Kappa measures the proportion of agreement between the test data and the modelled areas of species occurrence and absence (Fielding and Bell 1997). In presence-only modelling, AUC is the likelihood that a randomly selected presence point from test data is located at a site with a higher probability of species occurrence than that of a randomly selected point in the study area (Philips et al. 2006). Commission errors were calculated as the percentage of false positives in the test data, yielding a predicted distribution area of where the species in reality is absent (Araújo and Peterson 2012). In a similar way, omission errors were calculated as the percentage of false negatives in the test data.

To calculate these four parameters, we trained every distribution model with 75 % of randomly selected presence points whereby this model was cross-validated with test data in DIVA-GIS. Test data consisted of 25 % of the remaining observation data and pseudo-absence points (five times the number of presence points), randomly generated in the bounding box of the test data. Pseudo-absence points were restricted to this bounding box to reduce the number of such points that are located far away from the known, observed distribution range. This may inflate the values of the parameters (Lobo et al. 2008).

Finally, we tested with homogeneity χ^2 tests if application of consensus theory changes the rate of commission and omission errors according to experts.

3.3 Results

Expert evaluation

Of the 99 persons we invited to participate in the validation exercise, 45 responded. This yielded on average of almost nine experts per species. Experts came from 13 countries and were affiliated with universities, herbaria, and international, national, regional or non-governmental agricultural and environmental research institutions. One *B. excelsa* expert and one *C. odorata* expert were excluded from the analysis because they gave equal scores of one to all models proposed, i.e. that they considered them all being invalid. Although this gives us information about how relevant these models are for some experts in general, it does not give us information to discriminate between the models.

N. alpina experts reached the highest consensus between each other compared to experts from the other species. Therefore the variance explained by the first axis of the factor analysis was highest for their expert score correlation matrix (Table 3.1). For the other four species considerably less variance was explained by the first factor axis (Table 3.1).

Table 3.1 Variance in expert agreement explained in the first axis of the factor analysis.

Species	Number of experts	Variance explained in first factor
A. cherimola	9	0.29
B. gasipaes	5	0.39
B. excelsa	9	0.39
C. odorata	13	0.29
N. alpina	7	0.59

Quality and selection of distribution models

The median of consensus-weighed expert scores over all 45 species-variable combinations was 2.91, i.e. near to *Fair* according to the qualitative scores initially defined. These scores were higher than the corresponding un-weighed scores (Figure 3.1; Wilcoxon paired test, df = 44, $p = 0.049$). The median of un-weighed scores was 2.71.

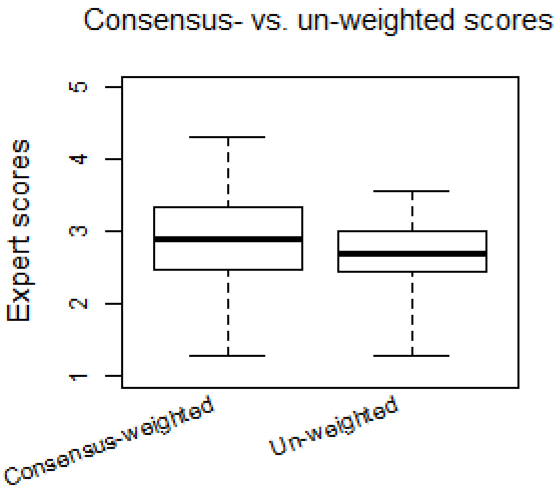


Figure 3.1 Boxplots of the averaged expert scores (n=45) when these are consensus-weighed and un-weighed

On average, variable combination 8 resulted per species in the best models according to un-weighed expert scores (Figure 3.2; Friedman, df = 8, $\chi^2 = 16.37$, $p = 0.04$), but according to consensus-weighed expert scores, no variable combination resulted in consistently better or worse models when taken in account all five species (Figure 3.2; Friedman, df = 8, weighed average expert scores: $\chi^2 = 14.05$, $p = 0.08$).

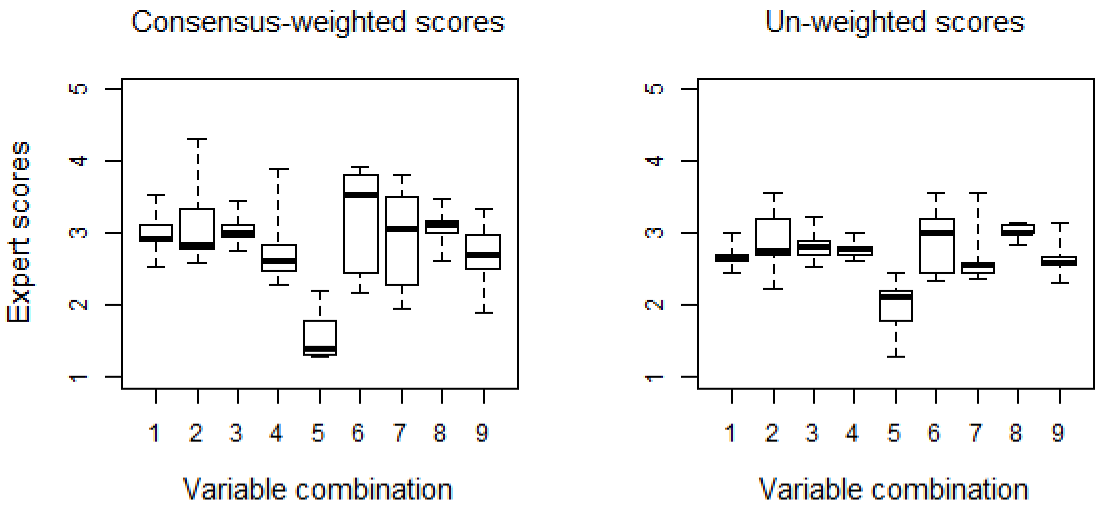


Figure 3.2 Boxplots per variable combination of averaged expert scores for each species (n=5)

The ranges between maximum and minimum consensus-weighed expert scores of the nine variable combinations per species were much higher compared to un-weighed scores (Appendix 3.3; Friedman, df = 8, $\chi^2 = 37.44$, $p < 0.001$). These wider ranges made it easier to select the best model per variable combination and per species compared to un-weighed scores (Figure 3.2; 3.3; Appendix 3.3).

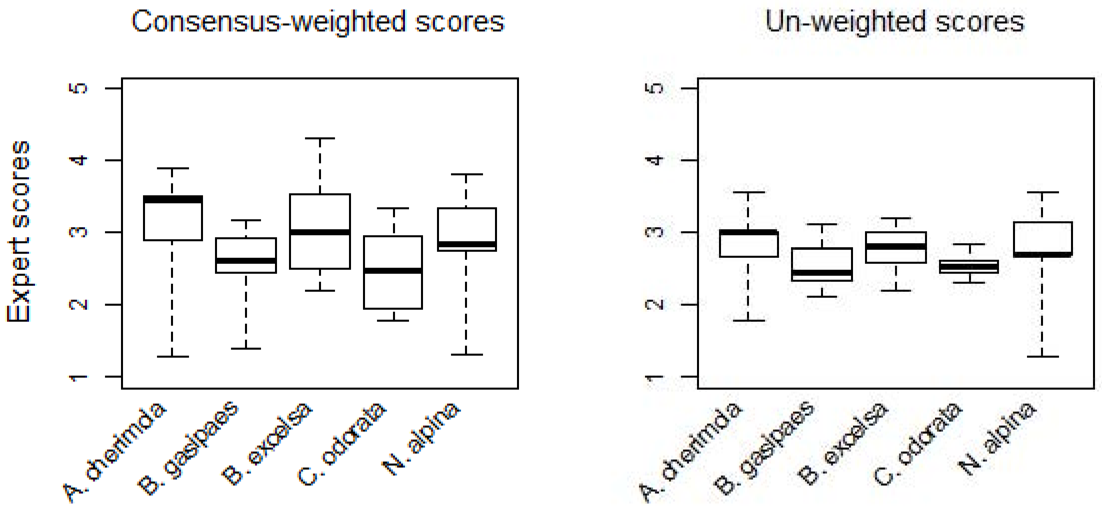


Figure 3.3 Boxplots per species of averaged expert scores for each variable combination (n=9)

For specific species, some variable combinations performed particularly well according to our consensus-weighed expert scores. The best *A. cherimola* and *N. alpina* models were close to the qualitative score *Good* (Figure 3.3; Appendix 3.3; respectively score 3.90 with variable combination 4 and 3.82 with variable combination 7). In the case of *B. excelsa*, the score of the best model was even between *Good* and *Excellent* (Figure 3.3; Appendix 3.3; score 4.30 with variable combination 2).

Correlation with model performance parameters

Both consensus-weighed and un-weighed specialist judgements resulted in significant correlations between the expert scores of all variable-species combinations and corresponding AUC, maximum Kappa and commission error (Table 3.3). Correlation between these parameters and un-weighed expert scores were similar to the correlation with un-weighed judgements of species specialists (Table 3.3). Expert opinions did not correlate significantly with omission error (Table 3.3). Almost all correlations with commission and omission errors were negative. This would be because expert appreciation and the rate of these errors are inversely related.

The best variable combinations according to the conventional parameters were different from the best model choice according to the experts independently if they were consensus-weighed or not. According to the AUC, maximum Kappa and commission error values, variable combination 4 resulted in the best distribution models (Appendix 3.3; Friedman AUC, df = 8, $\chi^2 = 25.63$, $p < 0.01$; Friedman Kappa, df = 8, $\chi^2 = 20.98$, $p < 0.01$; Friedman commission error, df = 8, $\chi^2 = 28.59$, $p < 0.0001$). The lowest omission errors were observed in variable combination 3 (Appendix 3.3; Friedman, df = 8, $\chi^2 = 15.73$, $p = 0.046$).

Considering each species individually, consensus-weighing only improved for *B. excelsa* the correlations between specialist judgments and the model performance parameters (Table 3.3). In the case of *A. cherimola*, we found highly significant correlations between the specialist evaluations and AUC, maximum Kappa and commission error (Table 3.3). For this species, similar results were obtained with weighed and un-weighed expert scores (Table 3.3). No clear correlations were observed for *N. alpina* and *C. odorata* (Table 3.3). Correlation between *B. gasipaes* expert scores and the model performance parameters worsened much when these scores were consensus-weighed (Table 3.3).

Table 3.3 Pearson correlation coefficients between expert scores and model performance parameters.

	Consensus-weighed expert scores			
	AUC	max Kappa	Commission error	Omission error
All experts (n = 43)	0.30*	0.37*	-0.33*	-0.07
A. cherimola (n = 9)	0.90**	0.83**	-0.90**	-0.40
B. gasipaes (n = 5)	0.19	0.11	-0.11	0.09
B. excelsa (n = 9)	0.87**	0.73*	-0.52	-0.43
C. odorata (n = 13)	0.23	-0.03	-0.22	0
N. alpina (n = 7)	0.20	0.58	-0.21	0.06
	Un-weighed expert scores			
	AUC	max Kappa	Commission error	Omission error
All experts (n = 43)	0.29*	0.39**	-0.31*	-0.05
A. cherimola (n = 9)	0.85**	0.86**	-0.83**	-0.38
B. gasipaes (n = 5)	0.54	0.42	-0.37	-0.18
B. excelsa (n = 9)	0.76*	0.65	-0.25	-0.23
C. odorata (n = 13)	0.13	0.06	-0.14	-0.08
N. alpina (n = 7)	0.27	0.59	-0.32	0.11

$*p < 0.05$, $**p < 0.01$

Commission and omission error according to experts

Averaged per species, 54 % of the preferred models had a commission error according to consensus-weighed expert judgment (Appendix 3.4). Forty-three percent of our species specialists indicated these were areas outside the natural distribution range. Twenty-two percent indicated that this was due to human disturbance such as selective extraction. Thirty-five percent did not specify the reason for species absence in predicted areas of occurrence (Appendix 3.4). For each species on average, 31 % of the experts indicated areas of species occurrence that were not predicted in his/her preferred model (omission) (Appendix 3.4). No significant differences were observed between the commission and omission errors according to consensus-weighed and un-weighed expert scores. Only a significant difference was observed between both values when we asked for the reasons of commission error (Homogeneity, df = 2, $\chi^2 = 10.80$, $p = 0.004$). The reason for this was that experts with higher competency values tended not to clarify the reasons for the existence of commission errors (Appendix 3.4).

3.4 Discussion

In this chapter, we present an approach using consensus theory to formalize expert knowledge to validate the outcomes of EEM. Consensus-weighed scores per species studied and per variable combination are on average higher and vary more than un-weighed model scores. This suggests that experts who have more favourable opinions about models reach more easily consensus between one another, whereas more sceptic experts do not converge to any consensus. Our results suggest that application of the consensus model could thus be a way to filter out sceptical ‘mavericks’ in the validation of models by experts.

The variation explained in the first axis of the factor analysis shows overall agreement between experts of a specific species group (Romney et al. 1986). This can be used as an indicator of the robustness of the model performance evaluation according to experts. Using this criterion, the expert selection of the best distribution model for *N. alpina* would seem to be the best because of the high degree of consensus on the quality of the modelling outcomes. However, the best model chosen by *N. alpina* experts didn’t coincide with the best model choice according to statistical analysis-based conventional model performance parameters. Similar discordance between expert evaluation and model performance after cross-validation with presence and pseudo-absence data were observed in another EEM outcome validation study as well (Anderson et al. 2003).

We suggest that in the case of high agreement between experts –such as in the case of *N. alpina*- their opinion should be considered seriously in the validation and selection of distribution models. In other cases, and when experts disagree, conventional parameters such AUC, Kappa, commission and omission error could be the lead parameters for model evaluation and selection.

N. alpina occurs in a restricted ecosystem in South America (temperate forests) that occurs only in two countries, i.e. Chile and Argentina. Consensus was much lower for the other four species that have a more extensive distribution range that covers three or more countries. Opposite opinions may arise because experts belong to geographical zones with differences in species niche occupancy, sampling density and environmental layer quality. On the basis of these differences, distribution models could be developed for different geographic zones and accordingly be evaluated separately by expert groups from these different geographic zones. In the software FloraMap, for example, it is possible to model species distributions for separate sub clusters of presence points located in different climate zones (Jones et al. 2002). This requires further research and expansion of expert validation exercises.

A significant correlation was observed between the 43 averaged expert scores and the corresponding Maximum Kappa, AUC, and commission error values. However, as mentioned before, experts and conventional parameters didn’t coincide in their best model choice. So, even though there is a significant relation between conventional parameters and expert evaluation, there are several discrepancies. The omission error values didn’t correlate significantly with expert scores. The calculation of omission errors may have been affected because presence point test data occur often relatively nearby presence point training data resulting in a spatial sorting bias (Hijmans 2012). Experts could know better where and to which extent species occur outside modelled distribution areas and thus could possibly better estimate omission error than statistical analysis-based parameters.

We didn’t find strong evidence that consensus-weighing improves the correlation between expert scores and AUC and Max Kappa. Only for one of the five species, i.e. *B. excelsa*, correlations between expert scores and conventional parameters clearly improved when these are consensus-weighed. Interestingly, this was also the species with the highest expert scores after consensus-weighing, with scores for the model with best variable combination (2) lying between the qualitative scores *Good* and *Excellent*. This model also had the lowest commission and omission errors after cross-validation (Appendix 3.3). However, we also found a significant decrease in correlation for *B. gasipaes* when expert scores were weighed. As it is, we only had very few *B. gasipaes* experts (n=5) compared to other species (n≥7). The low number of experts may explain why the consensus model didn’t perform well for *B. gasipaes* in reference to the conventional parameters.

According to our consensus-weighed expert scores average, model quality was towards *Fair,* whereas the best model choices per species, yield a value between *Fair* and *Excellent*. This indicates that these models are considered useful by our experts albeit their applicability remains limited in their opinion. In part, this may be explained because Maxent modelling could include areas in the model where the species is absent (commission error). This particularly affects model application for reserve design because areas may be included where the species is actually absent, which results in non-efficient investment in conservation measures (Araújo and Peterson 2012). For each species on average, more than half of our experts indicated that the model they considered best-performing, included areas where the species is absent.

Omission error was lower than the commission error. Thirty-one percent of the experts also indicated that areas of actual species presence were excluded by the model of their preference. The lower omission percentage suggests that these models are more appropriate for new population discovery and germplasm collecting than for reserve design.

Scale may also affect applicability of the modelled distributions (Guisan and Thuiller 2005). Maxent and other EEM software can predict the full distribution range of a species and can therefore be useful for evaluation at a national or regional scale. Many experts, however, tend to work at a more local scale and are only familiar with a part of the natural distribution range which they know in detail. On such a local scale, modelled natural species distributions tend to be less accurate than any expert’s knowledge of real field situation. In two cases, experts rated all nine potential natural distribution maps as invalid. This is most likely an indication that the modelled distributions were inaccurate (and thus not useful) at the local scale with which they were familiar. It is thus recommended to indicate to which scale distribution maps are accurate (Hurlbert and Jetz

2007; Lobo et al. 2008). However, EEM should also meet the needs for potential users. Therefore, more and more explanations are provided how to optimize Maxent and other EEM algorithms to respond to users' objectives (Elith et al. 2011; Araújo and Peterson 2012). Further improvement of techniques to model with a higher confidence local conditions would therefore be helpful to better their applicability.

According to our experts, 43 % of the commission errors in their preferred models, were predictions outside the species distribution range. Inclusion of spatial constraints in EEM may help reduce these over-predictions (Blach-Overgaard et al. 2010). According to the experts, 21 % of the identified commission errors in their preferred models, comes from the fact that species had become locally extinct due to selective extraction and forest degradation. It is a challenge to model these areas with EEM. To do this, absence points from these areas should be obtained, as well as demographic, geographical, or detailed remote-sensing layers that can help distinguishing areas of species occurrence from areas where the species is absent due to human disturbance.

We only asked experts if they observed commission and/or omission errors or not. In further studies, more details could be asked about the nature and extent of commission errors. However, a balance should be sought between depth of questioning and the ease for experts to respond.

Although no variable combination performed consistently better for all five species compared to other variable combinations according to consensus-weighed expert scores, it can be anticipated that a more optimal variable combination could further improve the results of EEM.

An important limitation in EEM is the lack of high resolution soil maps. Soil properties are known to be important factors for shaping the distribution of plant species (Coudun et al. 2006). Input of soil variables is especially relevant to precise modelled species distributions at landscape scale and departmental/province level (Pearson and Dawson 2003). However, currently only low-resolution soil maps are available at the regional level in Latin America and the Caribbean. The SOTERLAC soil map we used is still coarse compared to the interpolated bioclimatic layers that we used. Initiatives are underway to develop higher-resolution soil maps (Sanchez et al. 2009). Among other environmental variables that could improve model outcomes are solar radiation (Austin 2007) and Normalized Difference Vegetation Index (NDVI) (Prates-Clark et al. 2008).

It is clear that the results of EEM can also be improved by using better presence point quality and quantity (Anderson et al. 2003; Feely and Silman 2011). However, despite the fact that data points are increasingly shared by genebanks and herbaria through online portals such as GBIF, for many plant species only few presence points are available. Incomplete sampling and sorting bias is especially a problem when EEM is used to better understand species-environmental relationships (Elith et al. 2011).

Therefore there is an urge for more data collection in the field (Feely and Silman 2011). However field collection is expensive.

At the same time, incomplete sampling is also the main reason to use EEM in the case of predicting other areas where a species occurs naturally, on the basis of initial knowledge on its distribution. This is the principle use of EEM for *in situ* conservation planning and targeted collecting for herbaria and germplasm samples (Guarino et al. 2002). In this thesis, EEM has also been used considering this latter perspective.

Another, less-costly approach to improve the knowledge about species distributions is combining existing information obtained from experts with the results of EEM. This can be done e.g. by combining modelled natural species distributions with distribution range maps drawn by experts (Graham and Hijmans 2006) or correcting them based on existing descriptive literature (Rámirez-Villegas et al. 2010). This could be done in much more detail, when species specialists are directly involved in identifying the extent of natural species distributions and in the revision of presence point data. Especially relevant is local knowledge on species occurrence from under-sampled areas which are difficult to access for field inventories and germplasm collecting because of logistic and administrative constraints. Equally important, species specialists can also provide absence points (Tognelli et al. 2009). Both types of information enrich the understanding of species distribution and help to improve EEM as well.

Active involvement of existing national and international networks of foresters, taxonomists, ecologists, and/or nature conservationists could increase the number of participants in validation exercises. Amongst others, such networks are often established to facilitate sharing information. Indeed, several studies indicate that local experts are willing to share information on species occurrence. The clearest examples are with bird watching and reporting (Silvertown 2009), but there also cases where weed or other plant species are monitored (Silvertown 2009; Bradley and Marvin 2011). Such knowledge could be relevant for optimizing inventory programs that aim to minimize sampling biases (Feely and Silman 2011). It could also be used to iteratively improve the EEM to better predict species geographic distribution ranges and better understand species-environmental relationships.

3.5 Final summary

We obtained several interesting results about expert agreement, model appreciation and correlation of expert scores with conventional parameters. This confirms the potential of expert knowledge and the use of consensus theory for model validation. At the same time, we observed for several species low expert agreement and substantial discrepancies between expert scores and conventional parameters. We suggest that expert scores should be considered seriously when species specialists have reached high consensus. Consensus theory allows to increase the weight of the most knowledgeable experts in final model validation and to filter out sceptical ‘mavericks’. In the case of low expert agreement, however, conventional parameters may remain the leading reference to measure model performance. Low expert agreement may also be a result of geographically differences in model performance and expert knowledge domains. Further research should be carried out to better understand the possible occurrence of these zones and how to form geographically separate expert groups.

Online GIS and survey applications and the involvement of networks can facilitate the development of methods to carry out this type of consultation for large numbers of species, to interact in a time-effective way with many experts and present the generated natural species distribution maps for evaluation in an attractive and user-friendly way.

Case studies

4. Endemic wild potato (*Solanum* spp.) biodiversity status in Bolivia: reasons for conservation concerns[i]

Ximena Cadima[1,2], Maarten van Zonneveld[3,4], Xavier Scheldeman[4], Nora Castañeda[4], Fernando Patiño[1], Marcela Beltran and Patrick Van Damme[3,5,6]

[1]Fundación PROINPA, Casilla Postal 4285, Cochabamba, Bolivia
[2]Wageningen University, Biosystematics Group, PO Box 647, 6700 AP Wageningen, The Netherlands
[3]Ghent University, Faculty of Bioscience Engineering, Coupure links 653, 9000 Gent, Belgium
[4]Bioversity International, Regional Office for the Americas, PO Box 6713, Cali, Colombia
[5]World Agroforestry Centre (ICRAF), Global Research Programme 1, PO Box 30677, Nairobi, 00100, Kenya
[6]Institute of Tropics and Subtropics, Czech University of Life Sciences. Prague, Kamycka 129, Prague 6, Suchdol, 165 21, Czech Republic.

Abstract

Crop wild relatives possess important traits, therefore *ex situ* and *in situ* conservation efforts are essential to maintain sufficient options for future crop improvement. Bolivia is a centre of wild relative diversity for several crops, among them potato, which is an important staple worldwide and the principle food crop in this country. Despite their relevance for plant breeding, limited knowledge exists about their *in situ* conservation status. We used Geographic Information Systems (GIS) and Environmental Envelope Modelling (EEM) with the software Maxent to better understand geographic patterns of endemic wild potato diversity in Bolivia. In combination with threat layers, we assessed the conservation status of all endemic species, 21 in total. We carried out a complementary reserve selection to prioritize areas for *in situ* conservation and excluded 25 % of the most-threatened occurrence sites because costs to implement conservation measures at those locations may be too high compared to other areas. Following the IUCN Red List ecogeographic criteria Area of Occupancy (AOO) and Extent of Occurrence (EOO), at least 71 % (15 of 21 species) has a preliminary vulnerable status or worse. Our results show that four of these species require special conservation attention because they are highly threatened by human accessibility, fires and livestock activities pressure. Although highest species richness occurs in south-central Bolivia, i.e. in the

---

[i] *Submitted*

departments Santa Cruz and Chuquisaca, the first priority area for *in situ* conservation according to our reserve selection exercise is central Bolivia, Cochabamba, which is less threatened than the potato wild relatives' hotspot in south-central Bolivia. Only seven of the 21 species have been observed in protected areas. To improve coverage of potato wild relatives' distribution by protected areas, we recommend starting inventories in conservation parks and reserves with high modelled diversity. Finally, to improve *ex situ* conservation, we targeted areas for germplasm collecting trips of species with not any or less than five accessions conserved in genebanks.

4.1 Introduction

Crop wild relatives (CWR) include crop progenitors and their closely related species. Many of the latter species possess traits of interest for crop improvement, providing plant breeders with genes coding for biotic and abiotic stress resistance (e.g. resistance against pests and diseases, temperature, drought or salinity stress) or higher values for nutritional traits to name but a few (Tanksley and McCouch, 1997). Besides their role in providing genes for crop breeding, many CWR are already exploited by people in rural areas as they directly contribute to food security through provision of fruits, leaves, tubers and/or or seeds.

Most CWR are maintained *in situ* and their conservation status is often still largely unknown. Many CWR are increasingly menaced by habitat loss due to agricultural intensification, the impact of invasive species, deforestation, overgrazing and overexploitation (Maxted et al. 2008; VMABCC-BIOVERSITY 2009). In addition to these direct threats, global climate change is expected to become a long-term threat (Jarvis et al. 2008). However, immediate-term threats might require more urgent action to conserve sufficiently large populations that are resilient to the long-term threat of climate change.

In recent years, there has been a growing recognition of the global importance of CWR and the need for their conservation. The Convention on Biological Diversity (CBD 1992), the Status of Plant Genetic Resources for Food and Agriculture (FAO 2010a) and the Global Network for *In Situ* Conservation of Crop Wild Relatives (Maxted and Kell 2009), all highlight that active *in situ* (in wild populations and on farm) and *ex situ* conservation of CWR is essential for future crop improvement. Several global initiatives are currently being implemented to improve both *in situ* (VMABCC-BIOVERSITY 2009) and *ex situ* conservation (GCDT 2010) of CWR.

Bolivia is located in one of the main centres of origin of domesticated plants in the world (Vavilov 1951), and its high diversity of climatic conditions, soils and habitats combined with the high cultural wealth of indigenous peoples played a key role in the process of domestication (Ibisch and Mérida 2003). Bolivia is an important centre of diversity of several globally important staple crops such as potatoes (*Solanum* spp.), peanuts (*Arachis hypogaea* L.) and chili peppers (*Capsicum* spp.), but also crops of local importance such as the Andean grains, quinoa (*Chenopodium quinoa* Willd.) and amaranth (*Amaranthus* spp.), and Andean roots and tubers. Bolivia is also an important secondary centre of diversity of several other species such as maize (*Zea mays* L.), cassava (*Manihot esculenta* Crantz) and pineapple (*Ananas comosus* [L.] Merr.), and home to many wild relatives of all these crops.

Potato is production-wise the fourth most important crop in the world, after rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.) and maize. The crop and its wild relatives are therefore included in Annex I of the International Treaty on Plant Genetic Resources for Food and Agriculture, which facilitates the access to these genetic resources (http://www.planttreaty.org/texts_en.htm). In Bolivia, potato is the most important food crop for the local population with over 1000 native potato cultivars being cultivated by over 200,000 families (Zeballos et al. 2009; Cadima and Gandarillas 2009).

Despite the previously mentioned potential for breeding programmes, CWR are still underutilized in the development of new cultivars, albeit new technologies are available to better target their use (i.e. molecular maps, QTL analysis) (Hajjar and Hodgkin 2007). In the case of wild potato relatives (*Solanum* spp., section *Petota,* subsection *Potatoe*), several endemic Bolivian species have been studied, revealing traits important for future potato breeding (see Table 4.1). Ten species were found to show resistance against late blight (*Phytophthora infestans*), the main disease affecting potato production in Bolivia and elsewhere, while twelve species proved to be resistant to nematodes (*Globodera* spp.). Seven species show tolerance to abiotic stress, such as high temperature, drought or frost (Table 4.1; Hawkes and Hjerting 1989; Ochoa 1990; Spooner and Bamberg, 1994; Coleman, 2008; Jansky et al., 2008).

Table 4.1 Documented properties of endemic wild potato relatives of Bolivia.

Species	**Uses (Resistances)***	**References**
S. achacachense	Cyst nematode (*Globodera pallida*)	Hawkes y Hjerting, (1989) Ochoa, (1990) Centre for Genetic Resources (CGN), Netherlands Intergenebank Potato Database (USDA) International Potato Center (CIP), Peru Institute of Plant Genetic Resources and Crop Plant Research (IPK), Germany.
S. alandiae	Wart (*Synchytrium endobioticum*) Blackkeg (*Erwinia carotovora*) Cyst nematode (*Globodera pallida*) Flea beetle (*Epitrix cucumeris*), Potato aphid (*Macrosiphum euphorbiae*) Heat tolerance	
S. arnezii	Late blight (*Phytophthora infestans*) Blackkeg (*Erwinia carotovora*) Root-knot nematode (*Meloidogyne* spp.) Cyst nematode (*Globodera pallida*)	
S. avilesii	Late blight (*Phytophthora infestans*), Wart (*Synchytrium endobioticum*) Blackkeg (*Erwinia carotovora*) Root-knot nematode (*Meloidogyne* spp.), Cyst nematode (*Globodera pallida*)	

	Flea beetle (*Epitrix cucumeris*), Potato aphid (*Macrosiphum euphorbiae*).	
S. berthaultii	Late blight (*Phytophthora infestans*), Wart (*Synchytrium endobioticum*), Black scurf (*Rhizoctonia solani*), Verticillium wilt (*Verticillium* spp.) Blackkeg (*Erwinia carotovora*), Common scab (*Streptomyces scabies*), Bacterial wilt (*Ralstonia solanacearum*) Root-knot nematode (*Meloidogyne* spp.), Cyst nematode (*Globodera spp*.) Virus resistance PVX, PVY, PSTV Colorado beetle (*Leptinotarsa* spp.), Peach-potato aphid (*Myzus persicae*), Leaf hopper (*Empoasca fabae*), Flea beetle (*Epitrix* sp.), Leaf miner (*Liriomyza* spp.), Chinche (*Lygus* sp.), Spider mite (*Tetranychus* spp.)	
S. circaeifolium	Late blight (*Phytophthora infestans*). Blackkeg (*Erwinia carotovora*) Cyst nematode (*Globodera pallida*). Heat and Drought tolerance	
S. ×doddsii	Wart (*Synchytrium endobioticum*).	
S. flavoviridens	Peach-potato aphid (*Myzus persicae*), Colorado beetle (*Leptinotarsa* sp.), Spider mite (Tetranichus spp.), Leaf hopper (*Empoasca* sp.), Leaf miner (*Lyriomiza* spp.)	
S. ×litusinum	Late blight (*Phytophthora infestans*), Wart (*Synchytrium endobioticum*), Black scurf (*Rhizoctonia solani*) Cyst nematode (*Globodera* spp.) Colorado beetle (*Leptinotarsa* spp.), Chinche (*Lygus lineolaris*)	
S. neocardenasii	Peach-potato aphid (*Myzus persicae*, *Macrosiphum euphorbiae*), Leaf hopper (*Empoasca fabae*), Flea beetle (*Epitrix cucumeris*), Spider mite (*Tetranichus urticae*). Drought tolerance	
S. soestii	Late blight (*Phytophthora infestans*) Blackkeg (*Erwinia carotovora*) Cyst nematode (*Globodera* spp.) Heat tolerance.	
S. ugentii	Late blight (*Phytophthora infestans*) Cyst nematode (*Globodera* spp.)	
S. virgultorum	Late blight (*Phytophthora infestans*), Wart (*Synchytrium endobioticum*) Blackkeg (*Erwinia carotora*) Cyst nematode (*Globodera* spp.)	
S. gandarillasii	Drought tolerance Cyst nematode (*Globodera* spp.)	Same references plus Coleman (2008)
S. ×sucrense	*Verticillium* resistance Late blight (*Phytophthora infestans*), Wart (*Synchytrium endobioticum*). Cyst nematode (*Globodera* spp.)	Same references plus Spooner and Bamberg (1994)

	Blackkeg (*Erwinia carotovora*) Virus resistance PVX, PVA Potato tuber moth (*Phthorimaea operculella*) Frost resistance	
S. violaceimarmoratum	Colorado beetle (*Leptinotarsa* spp.) White mold (some) Late blight (*Phytophthora infestans*) Frost resistance	Same references plus Jansky *et al.*, (2008)

*Uses found (only) for 16 Bolivian wild potato species.

Wild potato relatives occur in the Americas from south-western United States to central Argentina and Chile. Some species, such as *Solanum acaule* Bitter, have a wide distribution range but most of them are confined to limited areas and ecological zones (Hijmans et al. 2002; Spooner and Salas 2006; Hawkes 1990). The wild potato species gene pool occupies an area wider than that of the native cultivated potatoes which are confined mainly to the South American Andes. Nevertheless, the highest number of wild potato relatives is also found in the Andes area from north-central Peru to central Bolivia (Hijmans et al. 2002).

In Bolivia, 35 wild species (following the classification of Spooner and Salas 2006) have been recorded, of which 21 species are endemic to the country (see Appendix 4.1). Wild potato species grow at altitudes between 700 to 4,500 m (Ochoa 1990) and occupy many different ecological niches in mesothermic and inter-Andean valleys, and in the subtropical Andean rainforest (Yungas). They are only absent from the Bolivian tropical lowland forests (Spooner et al. 1994).

Potato species can reproduce both sexually through insect-mediated pollination and asexually by means of stolons (e.g. runners) and tubers (Camadro et al. 2012). However, the role of these two reproduction strategies and dominance over another under different environmental conditions still needs to be determined for wild potato species (Camadro et al. 2012). Most potato species are allogamous (Salas et al. 2008; Camadro 2011). However, polyploidy species may have increased rates of autogamy (Camadro 2011). The latter species tend also to occur in more extreme climates (Hijmans et al. 2007). For example, the broadly distributed species *S. acaule,* occur at high altitude in a cold, harsh environment (Camadro 2011). This habitat lacks sufficient pollinators and the species reproduction thus relies on self-fertilization and asexual propagation (Camadro 2007).

In principle, potato species are annual. However, tubers of wild potato plants can persist for more than a year and resprout under favourable environmental conditions (pers. obs. X. Cadima). Natural hybridization between sympatric species can occur. From a breeding perspective there has been a lot of interesting on crossings between wild and cultivated species (Table 4.1; Camadro 2011). The probability of successful crossing between two species depends on their ploidy level and Endosperm Balance Number

(EBN) (Hijmans et al. 2002). EBNs are putative genetic factors that influence species crossing (Hijmans et al. 2002). Incompatibility of species with similar ploidy level is explained by differences in EBN. (Hijmans et al. 2002). Almost all wild potato species endemic to Bolivia are diploids except for *S. xsucrense*, *S. ugentii, S. hoopesii, S. bombicynum* (Appendix 4.2). These species are tetraploid (four sets of chromosomes, 4x) (Appendix 4.2).

There have been many efforts to collect germplasm of wild potato species in Bolivia. Nevertheless, a significant amount of the diversity remains unrepresented in genebank collections (Hijmans et al. 2000). For several species, only a few records exist whereas others are not conserved *ex situ* at all. At the same time, there is a limited knowledge about the *in situ* conservation status of these potato relatives (VMABCC-BIOVERSITY 2009). In February 2013, not one of the 21 endemic potato species has been listed yet in the online IUCN Red list (IUCN 2012).

Geographic information systems (GIS) are an effective tool that can contribute to generate new knowledge on and evaluate the conservation status of plant species (Brummitt et al. 2008). GIS are widely applied in different areas of environmental sciences and biodiversity, and have become an important tool in the development of strategies for the conservation and use of plant genetic resources (Jarvis et al. 2003). GIS are increasingly used to evaluate the geographic distribution and *in situ* conservation status of plant species, including CWR (Scheldeman et al. 2007; Penn et al. 2009; Hauptvogel et al. 2010; González-Orozco et al. 2012), as well as to guide targeted germplasm collecting trips (Jarvis et al. 2005; Scheldeman et al. 2007). Since species with a narrow distribution range are more prone to become extinct (Baillie et al. 2004; Işik 2011), spatial analysis has been widely used to assess species conservation status by identifying the extent of species distribution ranges (Willis et al. 2003). Spatial layers that contain information about human intervention (e.g. roads, agricultural conversion) can be overlaid in GIS over maps of species distribution and provide further information about the threats and conservation status of cultivated plant species and their relatives (Maxted et al 2008; Willemen et al. 2007) or ecosystems (Jarvis et al. 2010).

Recent collecting missions by PROINPA have increased the number of accessions for *ex situ* conservation (Patiño et al. 2008, Patiño and Cadima 2009). This new wild potato occurrence data combined with existing information about wild potatoes relatives' distribution and with new spatial information about threats allows a comprehensive survey of the conservation status of endemic potato wild relatives in Bolivia. In this study, we will (1) evaluate the *in situ* and *ex situ* conservation status of wild potato relatives based on spatial analysis; and (2) identify hotspots of endemic wild potato diversity, including areas that are threatened by human activities, causing disturbance to the habitat of the wild potato. The newly obtained results will all add to improve the conservation status efforts of several species and contribute to the maintenance of a future base for potato breeding.

4.2 Methods

Data sources

Georeferenced passport data from existing genebank databases (Centre for Genetic Resources of The Netherlands, United States Potato Genebank, Institute of Plant Genetics and Crop Plant Research of Germany, Intergenebank Potato Database and International Potato Center of Peru) were used to map the geographic coverage of the 21 Bolivian endemic wild potato species. Herbarium records on wild potato species developed by Hawkes and Hjerting (1989), Ochoa (1990) and Hijmans and Spooner (2001) were used to verify and improve the species distribution data. Duplicates were removed after merging the different data sets, and 331 georeferenced observation points remained. One hundred and one new presence points, obtained through PROINPA's germplasm collecting missions during 2006 to 2010 were added to this dataset. Additionally, 52 georeferenced herbarium and genebank records (presence points) were obtained from the Global Biodiversity Information Facility (GBIF). Twelve records from GBIF without coordinates were georeferenced based on locality descriptions with the use of Google Earth® and www.geonames.org, and were added to the analysis.

Species identification followed the taxonomy of Spooner and Salas (2006) which is commonly used in global databases and also in the Bolivian germplasm bank. We are aware that the results made in this study could eventually change if we take into account the last taxonomic treatment of wild potatoes reported in 2011 in the Solanaceae source website (www.solanaceaesource.org) that questions the delimitation between various species of the 'brevicaule complex' as defined by van den Berg et al. (1998) (Appendix 4.1).

The quality in taxonomic classification of observations points obtained from third parties such as through GBIF is often unknown (Chapman 2005). As an additional quality control, we therefore identified for each species, observation points in a-typical environments, which maybe potential erroneous observation points due to taxonomic misidentification. We calculated per species a-typical values using the 1.5 interquartile ranges as threshold for four bioclimatic variables that represent different facets of intra-annual climatic conditions: mean annual temperature, annual precipitation, temperature seasonality and precipitation seasonality. These calculations were done in R version 2.15.2 (R Development Core Team 2010). Climate values were obtained using the extract-values-by-point function of the R-based Raster package (Hijmans 2012) from the 2-5 minutes resolution Worldclim dataset (www.worldclim.org). We considered observation points as outliers when they scored a-typical values for two or more of the four climate variables. Following this method, only three observation points were identified as outliers of whom one came from the existing databases of herbarium records, and two from the recently collected herbarium and genebank records by

PROINPA. No outlier points were observed from the herbaria that made their data accessible through GBIF. These three ‘suspicious’ points all had a confident taxonomic classification and their coordinates were not only consistent at administrative unit level 1 (department level) but also at administrative unit level 2 (subregions in each department). We therefore decided to maintain these three points in the dataset.

Ecogeographic analysis

Ecogeographic analysis allows to identify for species putative ecotypes adapted to different environmental conditions, including rare and unusual ones. For our study, such an analysis would help to determine potentially interesting germplasm for potato breeding on adaptive traits to specific environmental conditions. We therefore identified for each endemic wild potato species its distribution across different climatic zones according to the climate classification of Köppen (see Kottek et al. 2006). Köppen was the first person to develop a quantitative global climate classification (Kottek et al. 2006). He did that as early as 1900 (Kottek et al. 2006). This classification is still used widely to distinguish climate zones (Kottek et al. 2006). It also has a plant ecological meaning. It has been developed on the hypothesis that because of differences in plant physiology, vegetation groups can be distinguished by climate zones (Kottek et al. 2006).

We used 30-seconds resolution monthly precipitation and mean temperature layers from the Worldclim dataset (www.worldclim.org) to define the different climate zones according to the criteria provided by Kottek et al. (2006). We did the calculation of these zones in R (R Development Core Team 2010). For the final map please refer to Appendix 4.3. In addition, we provide for each endemic wild potato species the altitudinal range in which they are occurring. Elevation data was derived from the 30-seconds resolution elevation data from the Worldclim set.

Species richness

A layer of the observed species richness based on presence points was created in DIVA-GIS using a five-minute resolution grid and applying a circular neighbourhood of 30-minute diameter (about 50 km around the equator) (see Scheldeman and van Zonneveld 2010). To estimate complete natural distribution ranges, we used an Environmental Envelope Modelling (EEM) approach. This technique defines the ecological niche, based on different environmental layers at the sites of the records, and identifies areas with similar environmental conditions as zones where the species could potentially occur and discriminates it from areas with an environment outside the ecological niche. Layers of the modelled areas of presence and absence for individual species can be stacked with the use of GIS to identify areas of potential species richness.

Maxent is a EEM tool (Phillips et al. 2006; Elith et al. 2011) for which the applied algorithm has been evaluated as performing very well, in comparison to other similar modelling software (Elith et al. 2006; Hernandez et al. 2006). Therefore, Maxent was selected to model the potential natural distributions of the 21 species. Nineteen bioclimatic variables representing different interannual bioclimatic conditions important for a plant's natural establishment and survival (Busby 1991), were used as environmental layers, derived from the Worldclim dataset and downscaled to a resolution of 30-seconds (~1 km) (Hijmans et al. 2005b). EEM with Maxent and these climate variables has been used successfully before to predict the occurrence of wild potato species (Simon et al. 2011). We therefore expect that this variable set will return also good quality modelling outcomes for the wild potato species in our study.

As a threshold to distinguish potential areas of occurrence from areas where a species would be absent, we chose the probability value where the value of sensitivity (true positive rate) plus specificity (true negative rate) is maximal. This is one of the thresholds that is recommended to generate EEM presence (grid cell value = 1) and absence (grid cell value = 0) maps (Liu et al. 2005). Then, to develop a potential richness map that is comparable with the observed richness map, we aggregated for each species, its presence-absence map to the same resolution as the observed richness map, i.e. five minutes. The aggregated cells received a value for species presence (grid cell value = 1) when species presence was modelled in one or more of its composing cells. Our final potential richness map consisted of the sum of all aggregated presence-absence maps.

Species with only few occurrence data may be sensitive to over-prediction in Maxent, although Maxent may even produce useful models with only 5-10 observations if these species have a rare and narrow distribution (Hernandez et al. 2006). This is likely true for several of our potato species that have a narrow distribution restricted to Bolivia: five of the 21 species had less than 10 unique locations (Table 4.2). Therefore, we restricted all generated potential distribution layers with a buffer zone around the Extent of Occurrence (EOO) to avoid overestimation of the modelled distribution ranges. A circular radius of 50 km was chosen for this buffer zone after the potato distribution maps developed by Hijmans and Spooner (2001). By restricting EEM with the buffer zones, our predictions of modelled species richness remain relative conservative.

In situ *conservation status*

As an indicator of *in situ* conservation status and on the basis of the presence points, we calculated for each species the EOO and the Area of Occupancy (AOO) (in km^2). These are categories in criterion B of IUCN red listing that indicate a taxon's vulnerability on the basis of its distribution range (IUCN 2010). These outcomes contribute to determine a preliminary conservation status. The two parameters were calculated on the basis of observed species distribution with the conservation assessment tools (CATS) extension in

ArcView 3.2 (Willis et al. 2003; Moat 2007). The CATS tool calculates the areas of AOO using the Equal Area Cylindrical Projection.

EOO is defined as the area within the shortest boundary that encompasses all occurrence sites. It is a measure of the distribution range in which a taxon occurs (IUCN 2010). Taxa with a higher EOO have a broader distribution range and are therefore less vulnerable to extinction compared to narrow-distributed taxa. AOO is a parameter that represents the area of suitable habitat for species occurrence within its EOO (IUCN 2010). This is calculated as the area of all grid cells in which one or more species records are located (IUCN 2010). The size of the grid cells can be calculated as a function of the extent of the distribution range (IUCN 2010). For each species, we chose after Willis et al. (2003) the 10 % of the maximum geographic distance between two collection sites to define the size of AOO grid cells. When a taxon has more suitable habitat within its EOO, it is less likely to go extinct within a defined time period.

The taxon must then meet at least two of three other options listed for criterion B to qualify for the vulnerable or worse conservation status (IUCN 2010). The options are (1) severely fragmented or known to exist in no more than a certain amount of locations; (2) continuing decline; and/or (3) extreme fluctuations in populations (IUCN 2010). However, this information requires intensive monitoring of specific populations, for which a substantial investment of funding would be needed. Alternatively, as a first indication for the amount of locations where the species occurs, we counted for each species the number of unique locations on the basis of our georeferenced species database.

As an additional parameter of *in situ* conservation, we calculated in ArcGIS 10 (ESRI, Redlands, California, USA), the number and percentages of records per species within protected areas. The protected area layer was derived from the World Database on Protected Areas (WDPA) (UNEP-WCMC 2010). All classes of protected areas were considered, i.e. national, international and private protected areas.

As an estimation of potential population decline, we used threat maps for natural ecosystems developed by Jarvis et al. (2010) to understand the major menaces for the endemic wild potato species and how these threats affect species richness. The layers consisted of six threats expected to occur within 2012 to 2015, i.e. accessibility to humans, conversion to agriculture, fires, livestock activities pressure, infrastructure, and oil and gas. The combined magnitude and sensitivity to threats was estimated for natural ecosystems that were defined by the Nature Conservancy (Jarvis et al. 2010). The magnitude of the different threats was calculated on the basis of existing datasets (Jarvis et al. 2010). Fire occurrence and frequency, for example, was detected using 250 m resolution MODIS satellite images (Jarvis et al. 2010). The spatial resolution of these maps was defined to 30-seconds (~1km) considering the precision of the various data sources and applicability for practitioners in the field (Jarvis et al. 2010). For more details on the sources of the datasets that were used to develop the threat maps and their availability please refer to Jarvis et al. (2010).

The sensitivity of each ecosystem to each threat was determined by expert consultation in workshops (Jarvis et al. 2010). Because the sensitivity has been determined at ecosystem level, the threat values should be interpreted with caution for specific species because some may be more sensitive than others to a specific threat. However, with the lack of species-specific information on threat sensitivity, we assumed that wild potato species populations from a specific habitat would have a similar level of sensitivity to the different threats as defined at ecosystem level.

Prioritization of areas for in situ *conservation*

We carried out a complementary analysis (Rebelo and Siegfried 1992) in DIVA-GIS (www.diva-gis.org), using a 30-minutes resolution grid (~50 km^2) to prioritize areas for *in situ* conservation. This analysis identifies the minimum number of grid cells required to conserve all species of interest. The grid cell with the highest number of species is being determined as the first priority area for *in situ* conservation. Second priority is given to the grid cell that covers the highest number of additional species that did not occur in the first priority cell. This prioritization exercise goes on until all species are covered by one or more cells.

We considered 30 minutes (~50 km^2) an appropriate scale to detect spatial patterns at country level. It is also a representative size for a protected area. The median size of the protected areas that are listed for Bolivia in the WDPA database is 36 km^2. The mean size of these registered conservation areas is 61 km^2.

Different approaches to define priority conservation areas were tested. In a first analysis, a complementary analysis was carried out without taking into account whether the locations of presence points are threatened or not. Secondly, only presence points at locations below the 75 percentile of average threat value were included in protected area selection (as highly threatened areas might be too costly to conserve). The reserve selection exercise was then repeated with only occurrence sites from protected areas. The latter analysis was carried out to evaluate how well the current protected area network in Bolivia conserves endemic potato wild relatives. This is the principal system for *in situ* conservation at national level. The representativeness of wild potato species in these conservation areas can therefore be considered an indicator for the conservation status of wild potatoes species. Finally, we carried out the reserve selection approach considering different putative ecotypes of each species that occur in the different Köppen climate zones.

Ex situ *conservation status*

To identify the *ex situ* conservation status of the 21 wild potato species endemic to Bolivia, we consulted the Global Strategy for the *Ex Situ* Conservation of Potato (van Soest 2006) which provides an overview of collected, and conserved, material in genebanks from Bolivia. We identified species not yet conserved in any genebank or with only a few accessions (less than five) conserved *ex situ*. We identified the areas where most of these species occur (gap analysis) on the basis of their occurrence sites targeting future collecting needs to improve the Bolivian wild potato species *ex situ* conservation status.

4.3 Results

Species richness

Wild potato relatives can be found from the northern high Andean part of Bolivia across the Andean-Amazon transition zone towards dry subtropical south-central Bolivia (Figure 4.1). Observed species richness is highest in south-central Bolivia (Figure 4.2), in Santa Cruz (inter-Andean valleys of Florida and Vallegrande provinces), and in Chuquisaca (provinces Zudañez, Azurduy Tomina and Oropeza). According to the potential species richness map, most species are expected to occur in northern Chuquisaca and Cochabamba (Figure 4.3). This area is situated more towards the centre of Bolivia than towards the inter-Andean valleys of Santa Cruz where currently most species are known to occur. The areas of high observed diversity are outside protected areas. The protected area where the highest amount of species is predicted to occur is ‘El Palmar’, but currently no wild potato species have been collected nor recorded from that area (Figures 4.2; 4.3). To a lower degree, the national park ‘Carrasco’ and ‘Tunari’ in Cochabamba and ‘Apolobamba’ in La Paz harbour endemic wild potato species (Figure 4.2; 4.3).

Ecogeographic analysis

Almost all species (20) and half of the specimens have been sampled in warm temperature climates with dry winters and warm summers according to the developed Köppen climate classification map (Table 4.2). In general, these areas correspond to inter-Andean valleys and mid-elevation subtropical forests. The second most-diverse climate zone is the cold arid steppe (Table 4.2). This zone is characterized by highland grass vegetation. With respect to breeding for adaptive traits for climate change adaptation such as for drought- and heat-tolerance and water-use efficiency, materials from the hot arid steppe climate are potentially interesting. This is the third-most rich and

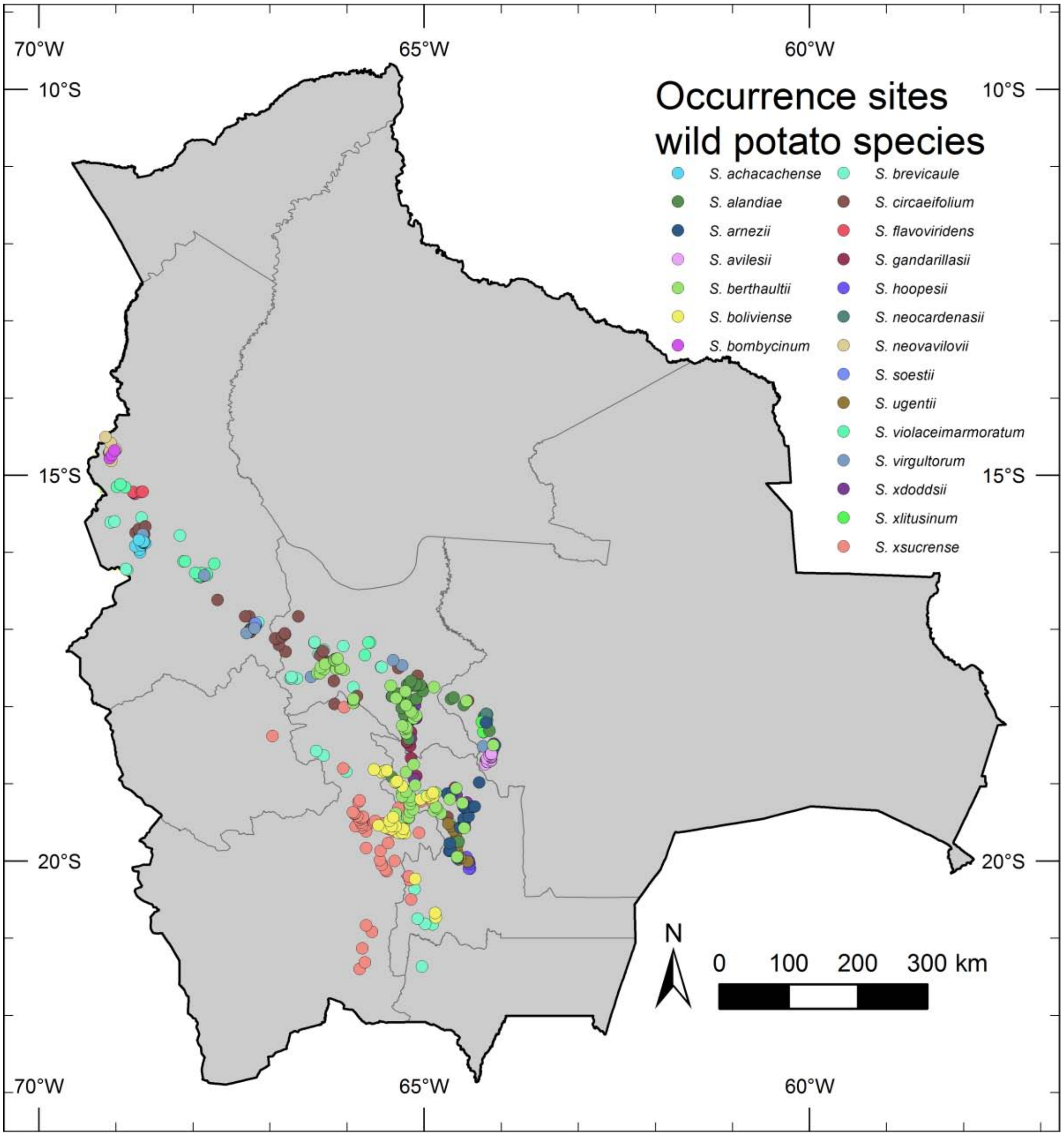


Figure 4.1 Distribution of the 21 endemic wild potato relatives on the basis of herbarium and genebank records.

abundant zone in endemic wild potato species (Table 4.2). Species occur above 1,200 masl (Table 4.2; Figure 4.4). It is common to find species above 3,000 masl (Table 4.2; Figure 4). Some species occur even up to elevations above 4000 masl (Table 4.2; Figure 4). Almost all species occur in two or more climate zones (Table 4.2).

Table 4.2 Distribution of species occurrence sites across Köppen climate zones and the altitude range in which they occur.

Species	As[1]	ET[2]	BWk[3]	BSh[4]	BSk[5]	Cfb[6]	Cwb[7]	Cwc[8]	min. alt. (masl)	max. alt. (masl)
S. achacachense		9					1		3,745	4,165
S. alandiae				9	3	1	21		1,633	3,377
S. arnezii				5	6		12	23	1,738	2,771
S. avilesii							19	19	2,145	2,841
S. berthaultii				16	20		35		1,692	3,219
S. boliviense					16		17		2,869	3,732
S. bombycinum		2					1		2,610	4,643
S. brevicaule		11		1	13		18	4	2,152	4,315
S. circaeifolium		3			2		36	1	1,933	4,753
S. flavoviridens	2						5	7	1,336	2,850
S. gandarillasii				19	1		1		1,411	2,740
S. hoopesii							11		2,360	3,950
S. neocardenasii				13	1				1,392	1,867
S. neovavilovii		1					16		2,444	4,155
S. soestii							6	6	2,862	3,595
S. ugentii							12	12	2,700	3,950
S. violaceimarmoratum	1					2	18	22	1,226	4,002
S. virgultorum		2		1			6		1,441	4,714
S. xdoddsii				5	8		5		1,977	2,762
S. xlitusinum				2	5		2		1,925	3,090
S. xsucrense		2	3	1	47		13		2,117	4,550
Total species richness	2	7	1	10	11	2	20	3		
Total observations	3	30	3	72	122	3	255	6		

[1]As = equatorial savannah with dry summer; [2]ET = tundra climate; [3]BWk = cold desert climate; [4]BSh = hot steppe climate; [5]BSk = cold steppe climate; [6]Cfb = warm temperature climate, fully humid and with warm summer; [7]Cwb = warm temperature climate with dry winter and warm summer; [8]Cwc = warm temperature with dry summer and cool summer.

Tentative IUCN conservation status

Following the preliminary IUCN red listing according to AOO (Area of Occupancy) or EOO (Extent of Occurrence) only, 24 % (five of the 21 species) of the endemic wild potato relatives is critically endangered (CR), which is due to their restricted observed distribution (Table 4.3). Another 19 % (four of the 21 species) is endangered (EN) according to these parameters, whereas 28 % (six of the 21 species) has a vulnerable status (VU) (Table 4.3). The remaining six species are not threatened (NT) or of low conservation concern (LC) based on the herbarium and genebank records (Table 4.3). To get a more complete assessment of the species conservation status following Red listing

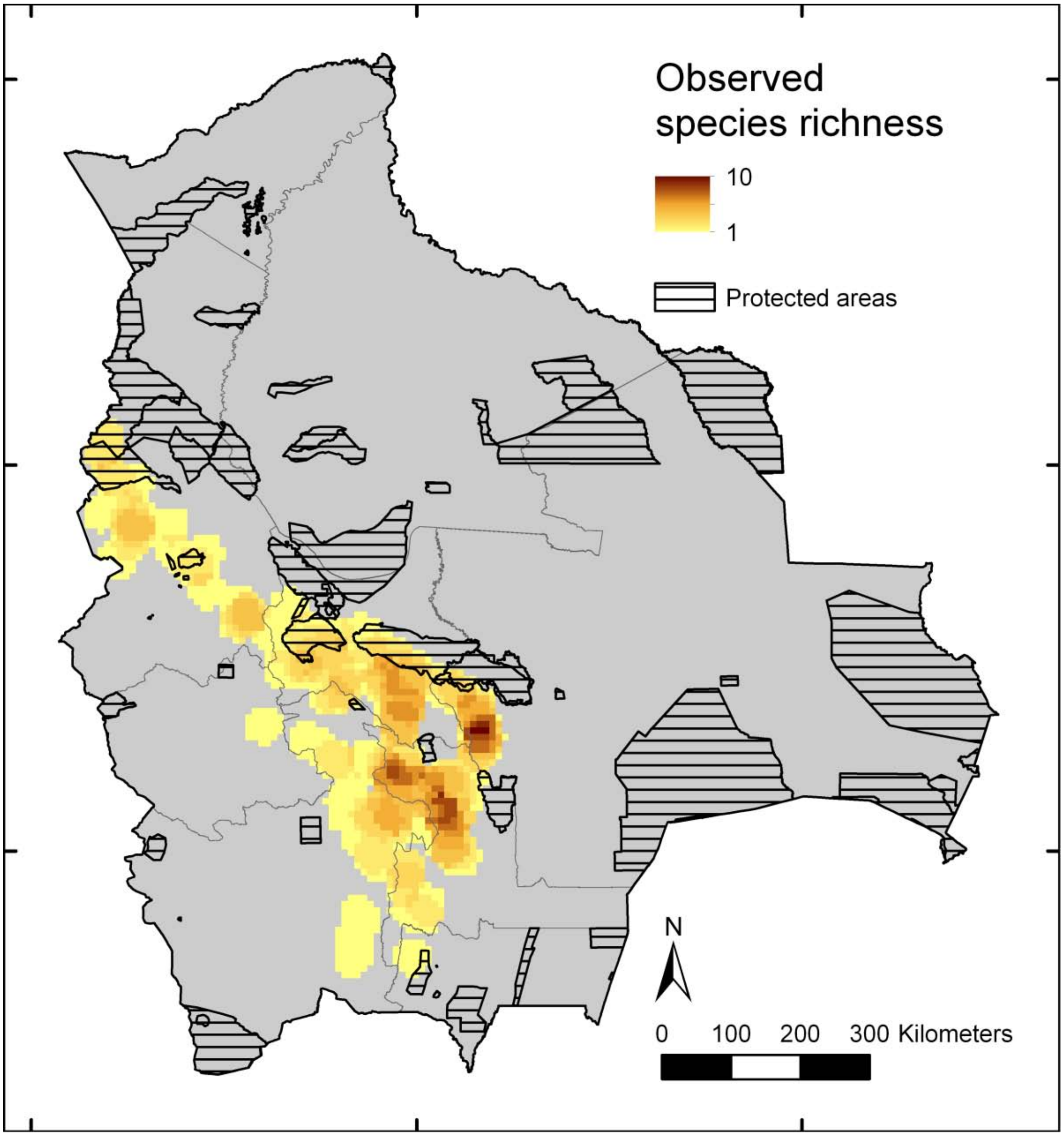


Figure 4.2 Observed wild potato species richness with a five-minute resolution grid-cell and 30-minute circular neighbourhood based the herbarium and genebank records of the 21 endemic wild potato relatives.

criterion B (IUCN 2010), we combined the AOO/EOO analysis with a threat assessment to identify which potato species require an IUCN conservation status of vulnerable or worse. According to our threat maps, the areas with highest average threat levels can be found in the western part of Cochabamba, and to a lower degree in northern Chuquisaca and western Santa Cruz where currently the highest numbers of species are observed (Figure 4.5). The most significant threats for all species considered in this study are

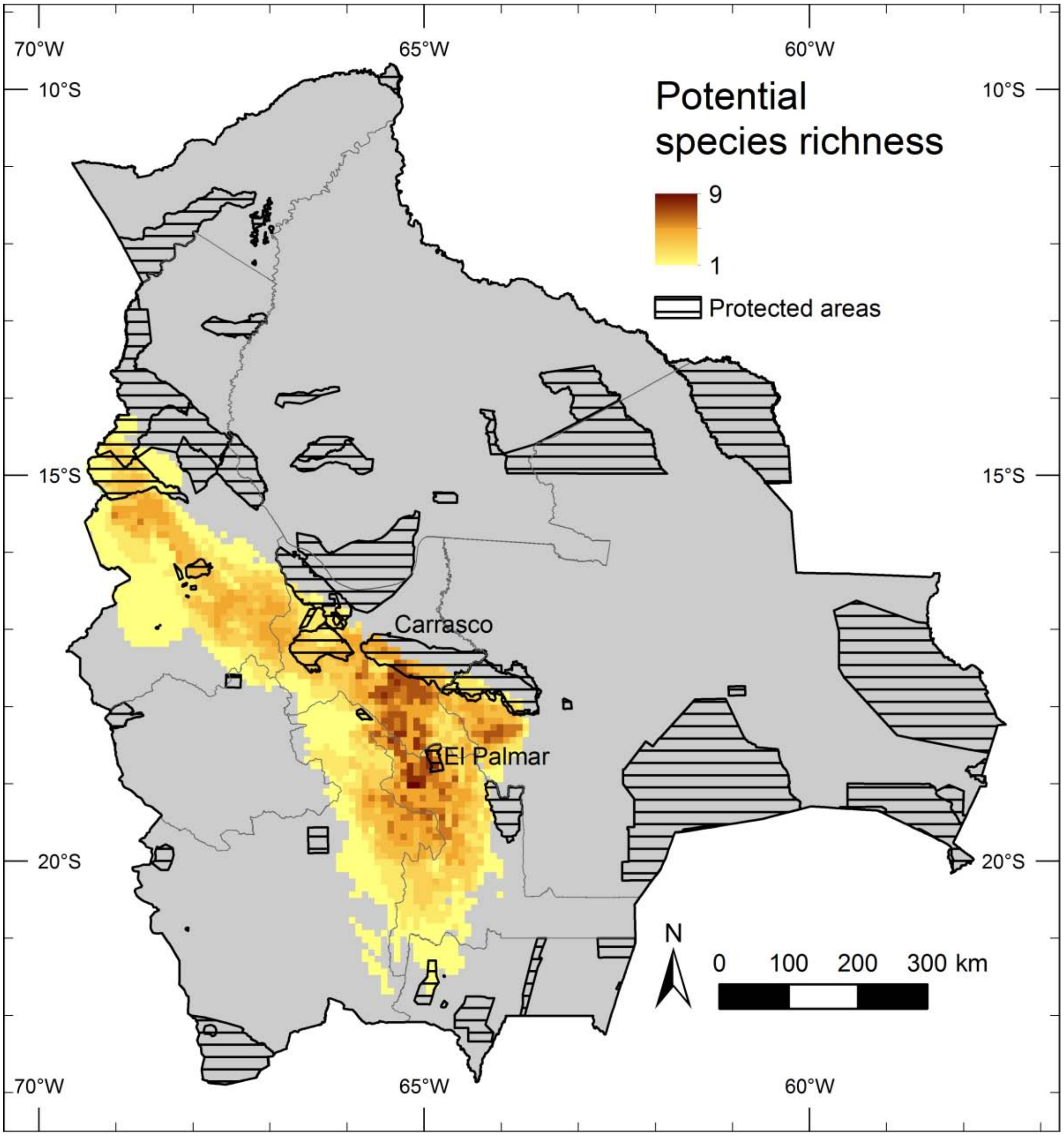


Figure 4.3 Potential wild potato species richness with a five-minute resolution grid-cell of the 21 endemic wild potato relatives using EEM in Maxent.

accessibility to humans, fire and to a lower degree livestock activities pressure (Table 4.3). A substantial part of the protected area ‘Tunari’ where potentially several potato species occur is being threatened as well by these pressures (Figure 4.5). Looking at the mean threat values, the seven most-threatened species are *S. achacachense* (EN), *S. arnezii* (VU), *S. brevicaule* (LC), *S. flavoviridens* (CR), *S. hoopesii* (EN), *S. ugentii* (EN) and *S. ×sucrence* (NT). Of these seven species, five species have a vulnerable conservation status or worse. Of these five species, *S. achacachense* has been observed in

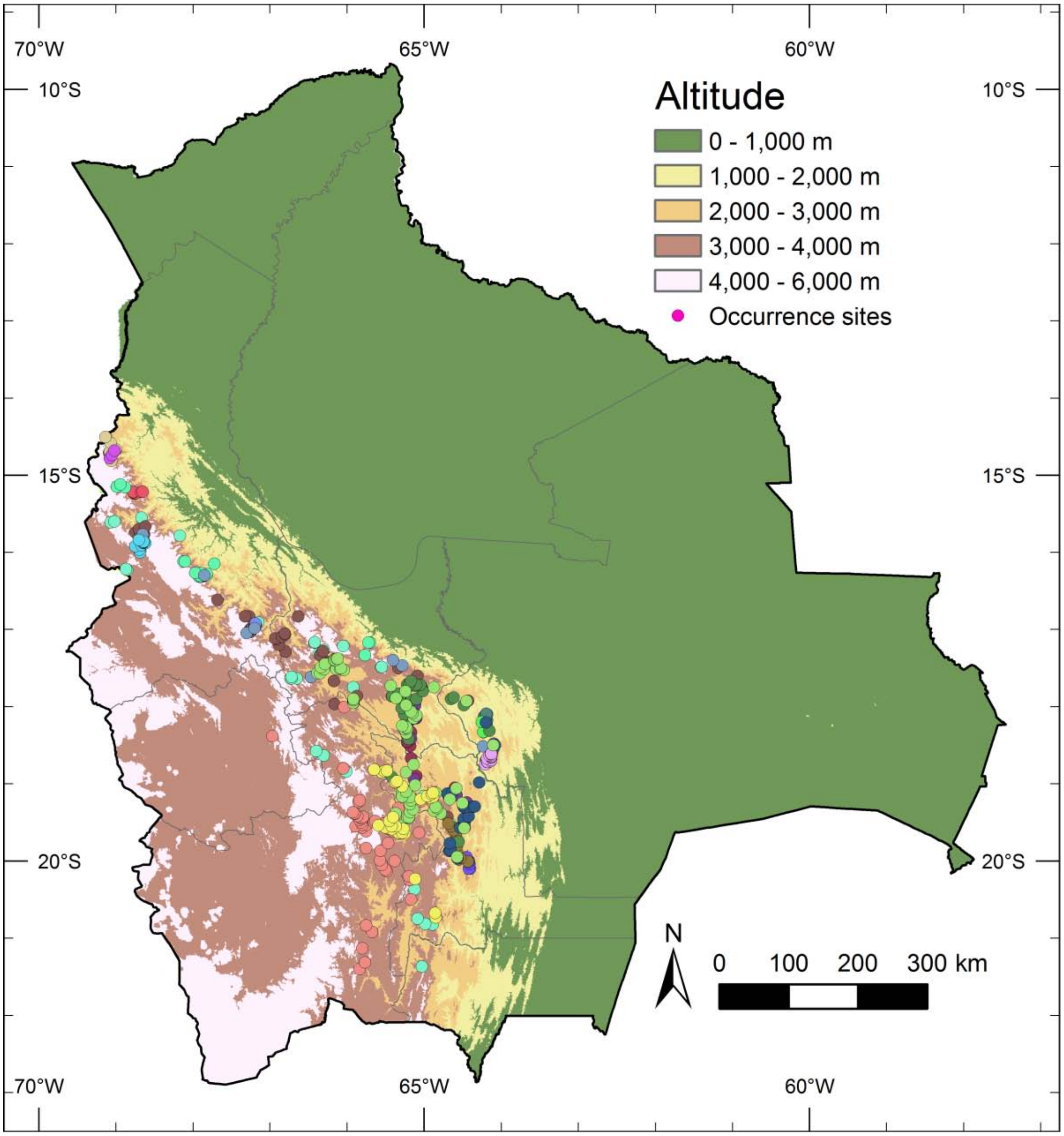


Figure 4.4 Distribution of endemic wild potato species across altitude ranges.

less than 10 unique locations and three species, *S. flavoviridens*, *S. hoopesii* and *S. ugentii*, in less than 15 locations (Table 4.3; Appendix 4.4).

Prioritization for in situ *conservation*

All 21 species can be conserved *in situ* in eight areas of ~50 km^2 when 25 % of the most threatened occurrence sites are not taken in account (Table 4.4). This is only one more area of ~50 km^2 than when all occurrence sites are considered in the prioritization of conservation areas, including those most-threatened.

Table 4.3 Total number of presence points of each endemic wild potatoes species in Bolivia, number of points in protected areas, preliminary IUCN conservation status, average threat value and identification of most important threats per species.

Species	Nr. of locations	Nr. of locations in protected areas	AOO (km^2)	EOO (km^2)	Tentative AOO/EEO Red listing status*	Mean threat value	Largest threat **	Second largest threat**
S. achacachense	10	0	29	129	EN	0.35	fires (0.86)	access (0.75)
S. alandiae	34	0	6874	20586	NT	0.30	access (0.60)	fires (0.53)
S. arnezii	23	0	5124	5488	VU	0.36	access (0.78)	livestock (0.71)
S. avilesii	19	0	38	59	CR	0.30	access (0.74)	Convers (0.61)
S. berthaultii	71	0	25085	36307	NT	0.30	access (0.84)	livestock (0.61)
S. boliviense	33	0	5205	10076	VU	0.29	access (0.95)	livestock (0.45)
S. bombycinum	3	3	5	0.3	CR	0.16	fires (0.39)	access (0.37)
S. brevicaule	47	13	111659	105673	LC	0.36	fires (1.05)	access (0.70)
S. circaeifolium	42	4	42095	46386	NT	0.27	fires (0.68)	access (0.56)
S. flavoviridens	7	4	39	67	CR	0.34	fires (0.95)	convers (0.46)
S. gandarillasii	21	0	2913	12308	VU	0.27	access (0.68)	livestock (0.49)
S. hoopesii	11	0	264	430	EN	0.34	fires (1.00)	livestock (0.57)
S. neocardenasii	14	0	37	507	CR	0.28	access (0.75)	fires (0.56)
S. neovavilovii	17	17	61	180	EN	0.17	fires (0.52)	access (0.35)
S. soestii	6	0	1	3	CR	0.16	access (0.57)	livestock (0.29)
S. ugentii	12	0	324	401.4	EN	0.42	fires (1.28)	livestock (0.60)
S. violaceimar Moratum	22	9	8830	13703	VU	0.28	fires (0.73)	access (0.65)
S. virgultorum	9	2	18792	25035	NT	0.18	access (0.63)	livestock (0.22)
S. ×doddsii	18	0	3268	11985	VU	0.20	access (0.65)	livestock (0.55)
S. ×litusinum	9	0	1663	10161	VU	0.29	access (0.80)	livestock (0.58)
S. ×sucrense	66	0	25436	48284	NT	0.37	fires (0.99)	access (0.86)

*CR: Critically Endangered; EN: Endangered; VU: Vulnerable; LC: Lower Concern; NT: Not threatened.
** access: accessibility to humans; livestock: livestock activities pressure; convers: conversion to agriculture.

Table 4.4 Results of reserve selection analysis to prioritize areas for *in situ* conservation.

Methodology	Nr. of cells included	Nr. of species included
All occurrence sites are included in the reserve selection (threats not taken into account)	7	21
25 % of the occurrence sites with the highest average overall threat not included in the reserve selection	8	21
Only occurrence sites protected areas are included in the reserve selection	3	7

By excluding 25 % of the most-threatened occurrence sites, the areas of highest species richness, i.e. in northern Chuquisaca and western Santa Cruz, were less taken in account in the reserve selection because large parts of natural vegetation in those areas are threatened by humans due to increased accessibility, fire occurrence and livestock activities pressure (Figure 4.6). Instead, the area of highest priority is in south-eastern Cochabamba, where six species can be conserved *in situ* in an area of 50 km^2. The second priority are the northern highlands in western La Paz where three additional species can be conserved in an area of 50 km^2, which moreover is within a protected area (Area Natural de Manejo Integrado de Apolobamba [Figure 4.6]). The third priority area for conservation is western Santa Cruz were two additional species could be conserved. The fourth priority area is located in La Paz too. The latter prioritized area also covers the only observed locations of the endangered species *S. achacachense* (Figure 4.6, Appendix 4.4). The endangered and highly threatened species *S. hoopesii* and *S. ugentii* are located in Chuquisaca (Figure 4.6, Appendix 4.4). When we restricted the reserve selection to only the protected areas, only seven (33 %) of the 21 species could be conserved and of the four most endangered species only *S. flavoviridens* was included.

Ex situ *conservation*

According to data reported in the Global Strategy for the *Ex Situ* Conservation of Potato (van Soest, 2006) updated with data from PROINPA, there are 10 genebanks in the world holding 1062 accessions of the 21 wild potato species endemic to Bolivia (Appendix 4.5). This may include duplicates of exchanged materials. The *ex situ* collection in Bolivia maintained in the National Genebank of Andean tubers and roots is the result of repatriated materials from the Centre of Genetic Resources the Netherlands (CGN) and new collecting trips in recent years. This national collection has currently 235 accessions of 18 endemic species (the total potato wild collection has 618 accessions, including other non-endemic species occurring in Bolivia). Sixty-five of these, concern new material collected over the 2006 to 2010 period.

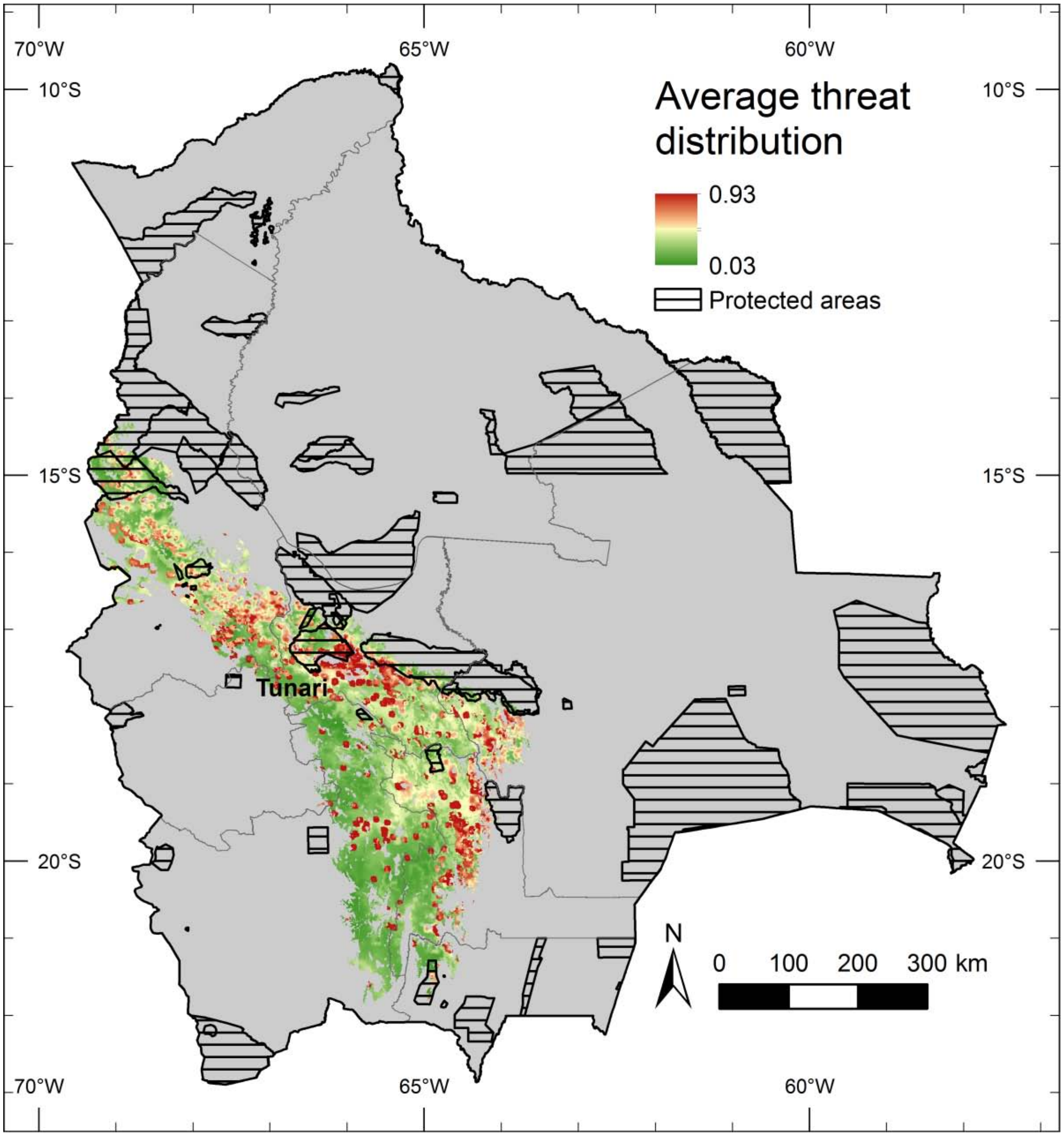


Figure 4.5 Mean threat values (average of human accessibility, conversion to agriculture, fires, livestock activities pressure, infrastructure, and oil and gas) in a 30-second resolution map across the modelled distribution range of endemic wild potato species in Bolivia.

Some species are well-represented in the genebank collections, such as *S. berthaultii* which has the largest number of accessions (228), followed by *S. ×sucrence* (195) and then *S. boliviense* (141). On the other hand, no germplasm of *S. bombycinum* and *S. ×litusinum* is conserved in any *ex situ* collection. Other species poorly conserved are *S. neovavilovii* (two accessions), *S. soestii* (two) and *S. flavoviridens* (four). Samples of these species only exist in the Bolivian collection (Appendix 4.5). The small number of

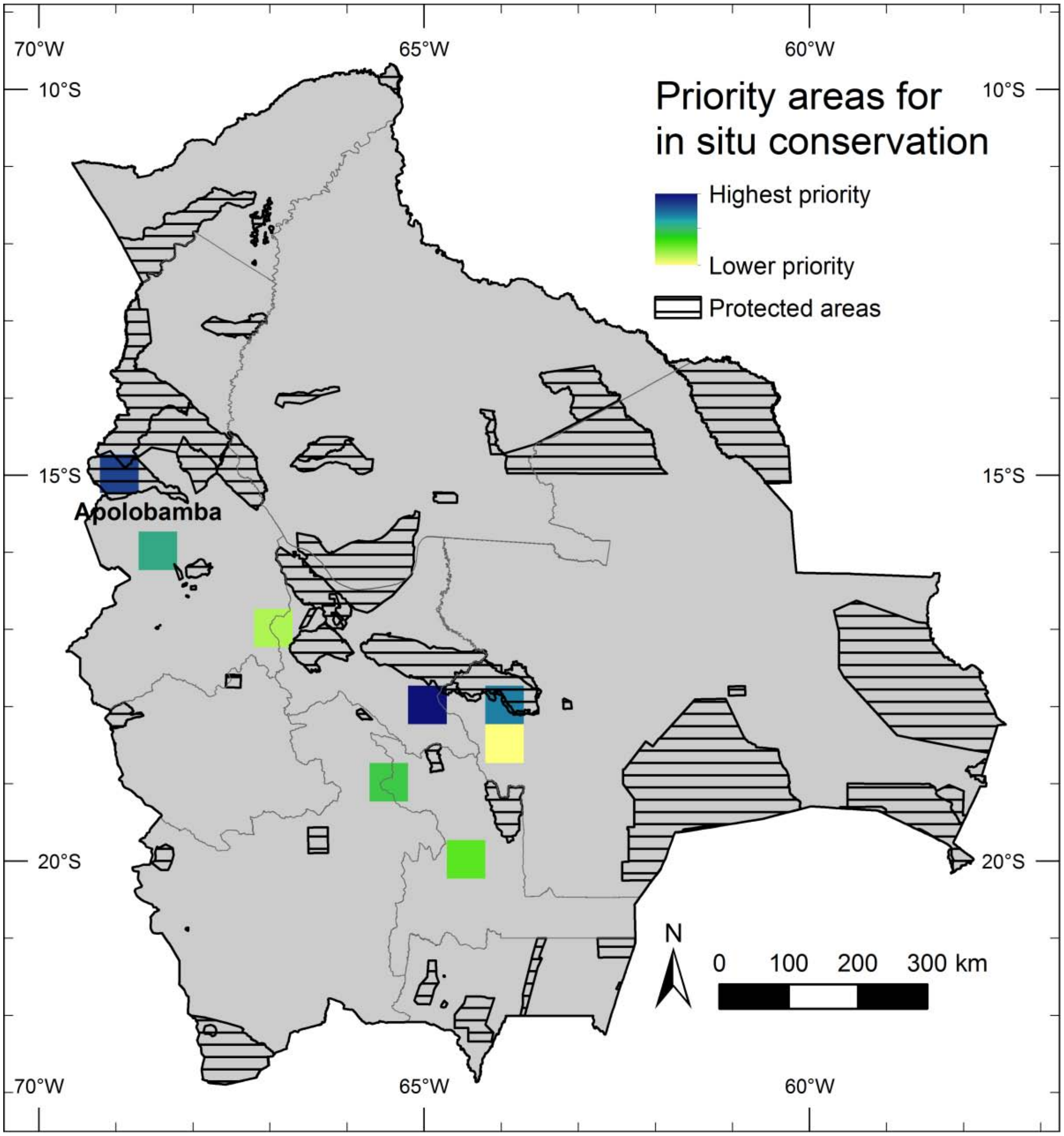


Figure 4.6 Prioritized areas to conserve *in situ* 21 endemic wild potato species with the use of the complementary reserve selection and excluding 25 % of the most threatened locations where the species have been collected or recorded.

samples for these species in genebanks also coincides with a restricted distribution in the field and limited accessibility to reach the natural habitats of occurrence of these species. Prioritized areas for collecting trips are La Paz (Provinces Tamayo and Saavedra) where populations of *S. flavoviridens, S. neovavilovii* and *S. bombycinum* have been observed (Figure 4.7). *S. soestii* could be explored in La Paz (Province Inquisivi) and Cochabamba (Province Ayopaya). *S.* ×*litusinum* is most likely to occur in the Cochabamba-Santa Cruz border area and at the frontier between Potosi and Chuquisaca (Figure 4.7).

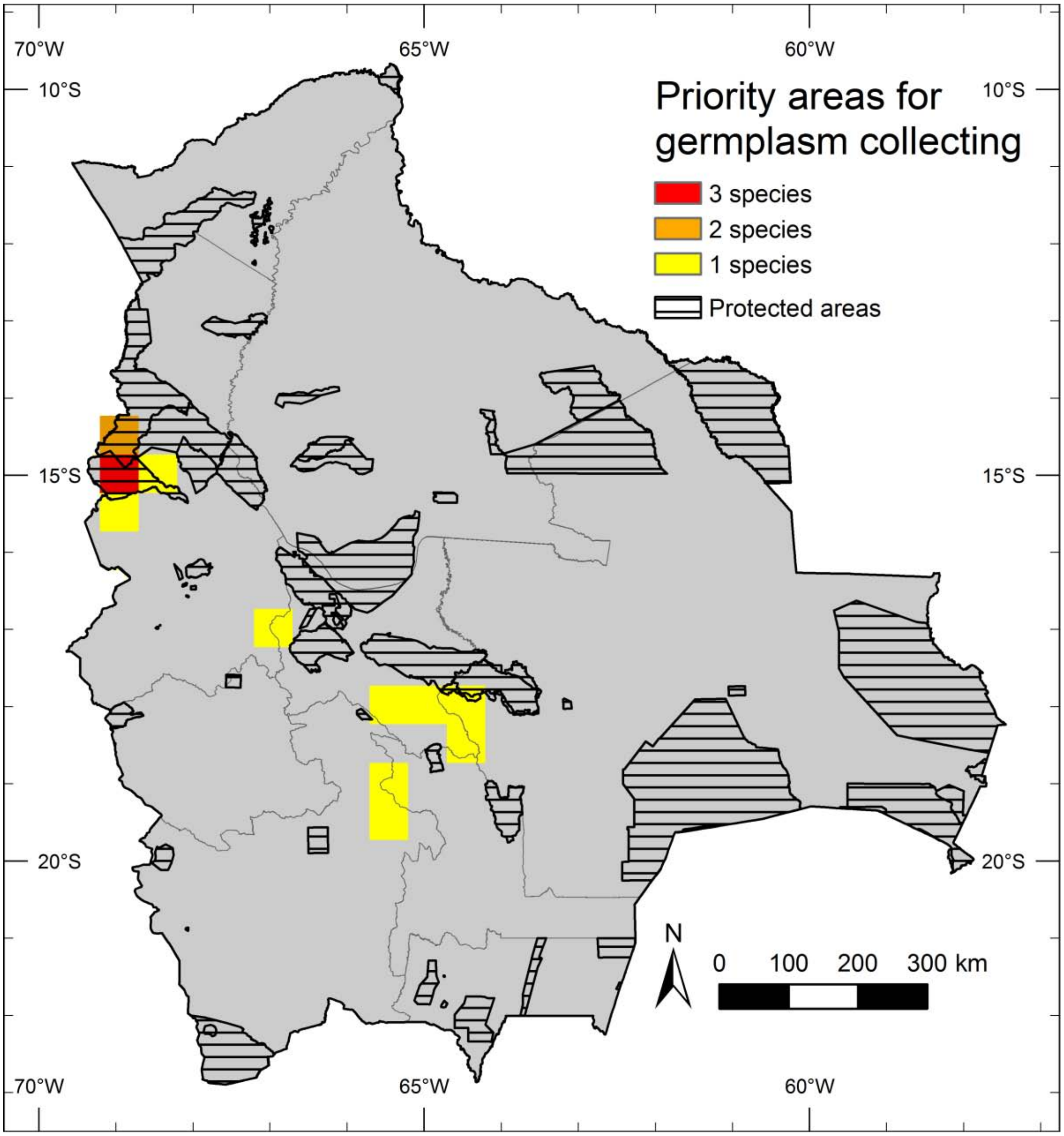


Figure 4.7 Map with prioritized cells to target germplasm collecting trips of the five potato wild relatives for which no or less than five accessions are currently conserved (*S. bombycinum, S. ×litusinum S. neovavilovii, S. soestii* and *S. flavoviridens*).

Comparison of conservation priorities of species and putative ecotype diversity

In addition to a reserve selection exercise at species level, we also carried out a prioritization of areas for conservation considering the different putative ecotypes according to the climate zones. Ecotypes within wild potato species of interest can be useful for breeding on adaptive traits related to specific climate conditions. In total, we identified 56 putative ecotypes for the 21 endemic wild potato species (Table 4.2). Fortynine of these possible ecotypes occur in the 75 % least threatened collection sites (Table 4.6). These are scattered across the wild potato distribution range in Bolivia and can be captured in 19 grid cells (Table 4.6).

When we excluded 25 % of the most-threatened collection sites, the area with highest ecotype diversity coincides with the one of highest species diversity (Figure 4.8). And also other areas of unique putative ecotype diversity coincide with areas of high species diversity such as the northern highlands in western La Paz (Figure 4.8). In addition to the targeted areas for 'species' *in situ* conservation, a new prioritized area of unique high putative ecotype diversity is observed in eastern Potosí in the climate zone 'Cwb', i.e. with warm temperature climate, dry winter and hot summer (Figure 4.8).

Table 4.6 Results of reserve selection analysis to prioritize areas for *in situ* conservation.

Methodology	Nr. of cells Included	Nr. of putative ecotypes included
All occurrence sites are included in the reserve selection (threats not taken into account)	20	56
25 % of the occurrence sites with the highest average overall threat not included in the reserve selection	19	49
Only occurrence sites protected areas are included in the reserve selection	7	12

Seven of the 56 putative ecotypes occur exclusively in the 25 % of the most-threatened collection sites. These are *S. circaeifolium*, *S. gandarillasii*, *S. neocardenasii* populations in cold arid steppe climate; *S. virgultorum*, *S. xsucrense* populations in hot arid steppe climate; *S. neovavilovii* populations in tundra climate; and *S. violaceimarmoratum* in equatorial savannah with dry summers. These seven putative ecotypes were represented by only one occurrence site and are therefore likely to be species populations in extreme environments. All these endangered ecotypes are already conserved *ex situ* but these putative ecotypes should be targeted for further germplasm collecting as they may be susceptible to *in situ* extinction (Figure 4.9).

4.4 Discussion

Some 71 % (15 species) of the endemic wild potato relatives has a preliminary vulnerable or worse status according to the IUCN criterion B category AOO (Area of Occupancy) and EOO (Extent of Occurrence). Of these, five species are of particular concern for protection because they are facing significant threats, particularly by fire (*S. achacachense*, *S. arnezii*, *S. flavoviridens*, *S. hoopesii* and *S. ugentii*) (Table 4.3).

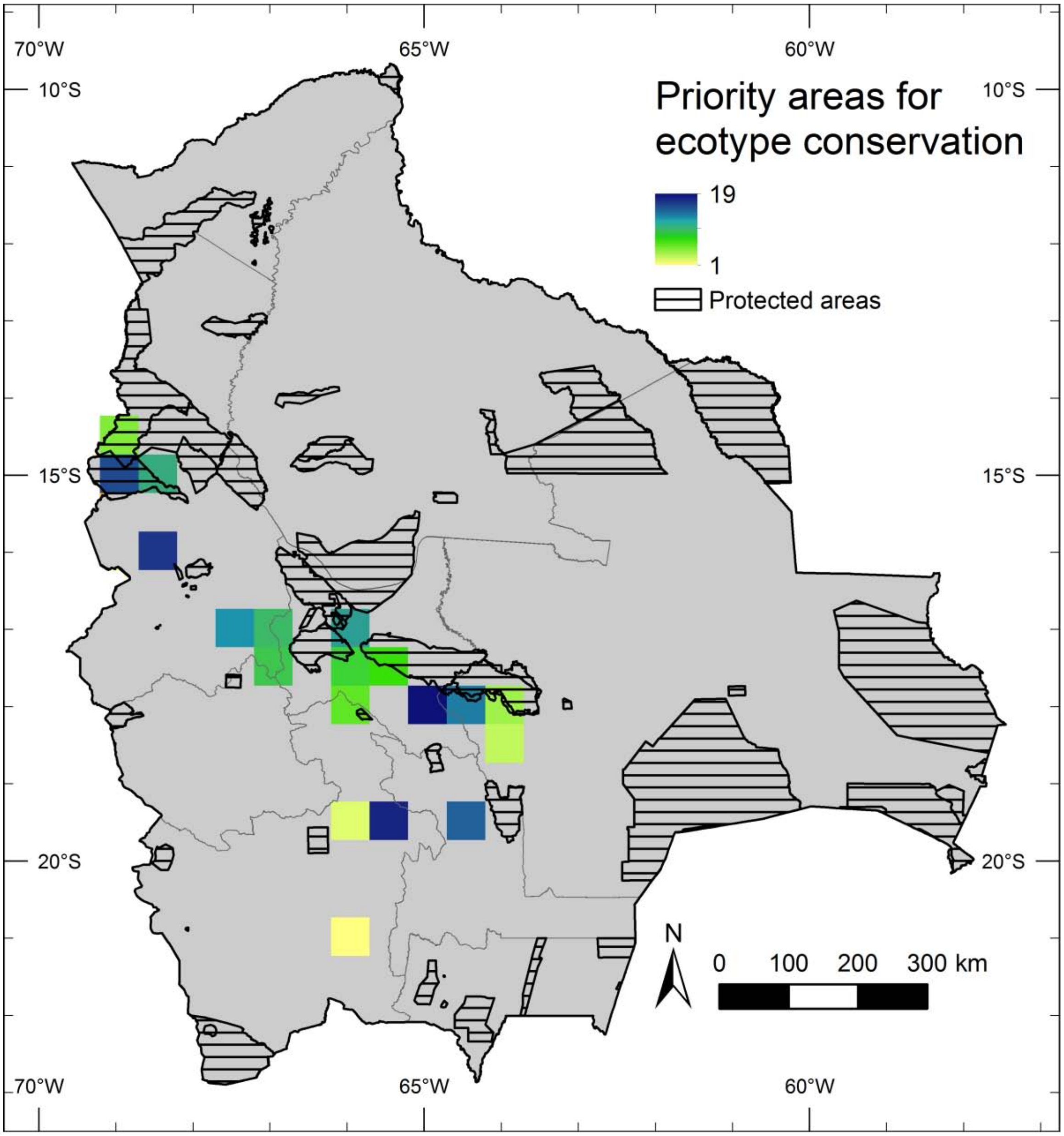


Figure 4.8 Prioritized areas to conserve *in situ* the 56 putative ecotype of the 21 endemic wild potato species with the use of the complementary reserve selection and excluding 25 % of the most threatened locations where the species have been collected or recorded.

Of these five species, *S. achacachense, S. flavoviridens, S. hoopesii* and *S. ugentii* have only been observed in a restricted number of locations (<15). These four species qualify most for a conservation status of vulnerable or worse according to criterion B of the IUCN red listing assessment and should therefore be prioritized for conservation.

Among these species, *S. flavoviridens* is underrepresented in genebanks. Of the other species, fortunately a considerable number of accessions is conserved *ex situ*.

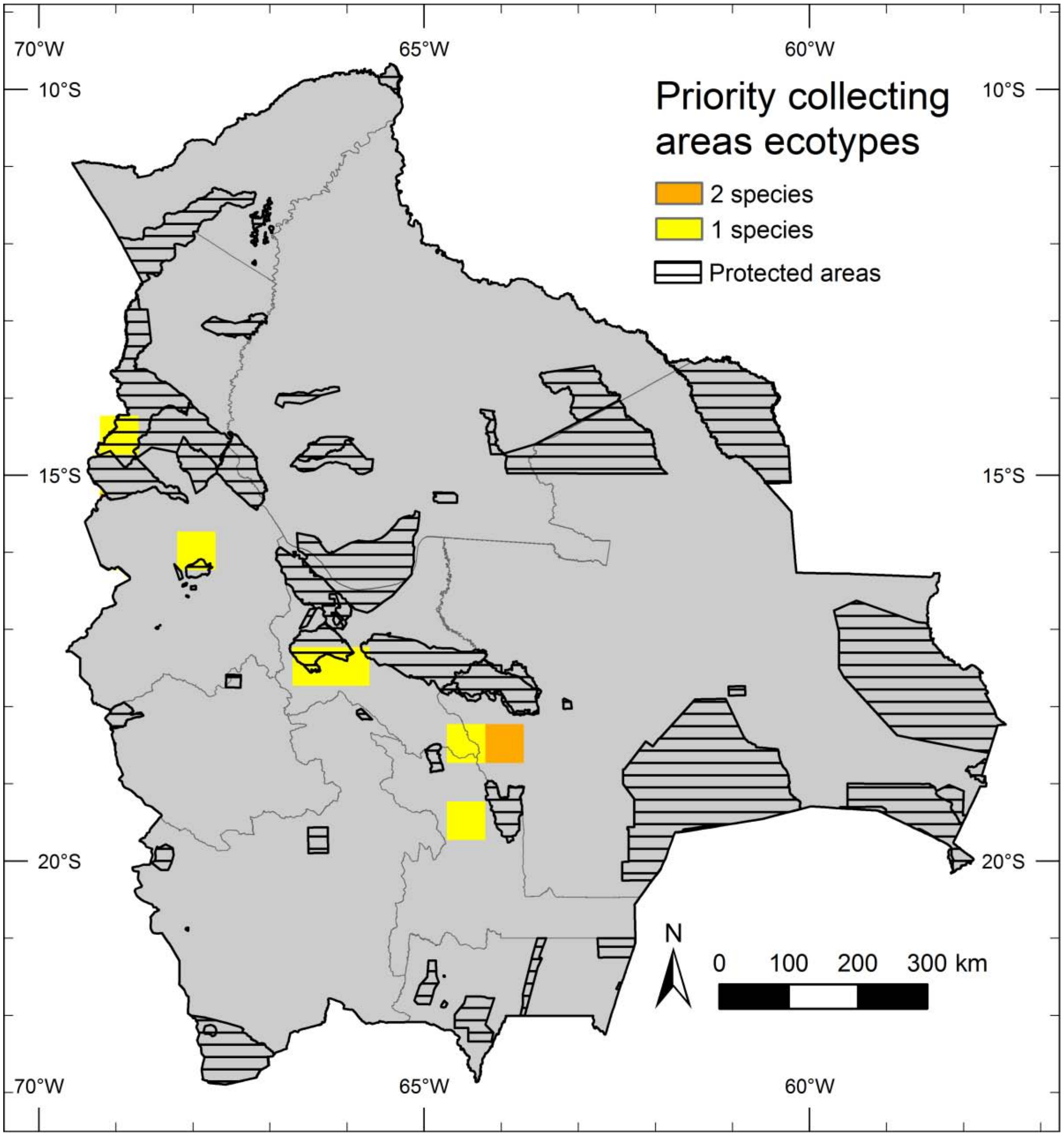


Figure 4.9 Map with prioritized cells to target germplasm collecting trips of the seven endangered putative ecotypes that occur exclusively in the 25 % most threatened collection sites.

Occurrence sites of *S. brevicaule* and *S.* ×*sucrense*, i.e. two species that have a broader distribution than the five species mentioned above, are also highly threatened. Although these threats may have a substantial impact on the genetic diversity of the populations of these two species, new occurrence sites in less-threatened parts of their distribution range may be identified for their *in situ* conservation at species level. Species distribution modelling can help in identifying those areas.

In addition to *S. flavoviridens,* four other species should be prioritized for targeted collecting trips because they are either not yet conserved in any genebank (*S. bombycinum* and *S.* ×*litusinum*) or are underrepresented (*S. neovavilovii* and *S. soestii*) (Appendix 4.5). The department of highest priority for collecting is La Paz (Provinces F. Tamayo and B. Saavedra) within the protected area 'Area Natural de Manejo Integrado de Apolobamba' where three of the five prioritized species for collection have been documented to occur. The areas in the north-western part of La Paz are locations of difficult access which would explain the few samples collected in these areas.

The scenario in which we excluded 25 % of the threatened occurrence sites had our preference to prioritize areas for *in situ* conservation because the reduced cost of *in situ* conservation in less-threatened areas may outweigh the cost of implementing conservation measures in an additional area. However, all priority areas identified for conservation, except one that lies within a protected area in northern La Paz (Apolobamba), are areas where farming is important. In the case of Santa Cruz, livestock is also important. These areas are not related to any system of conservation or protection, so even while we excluded 25 %t of the most-threatened sites, the other locations may still be vulnerable to threats as a result of human activities. For example, although *S. virgultorum* occurrence sites do not have particular high threat values, known populations of these species reported in the past (Ochoa 1990) were not found back in recent field visits (between 2006 and 2010). Similar indications of decline may even be more pronounced in populations of species that are highly threatened according to our analysis.

Studies on the effectiveness of conservation efforts of vertebrates to reduce their threat level demonstrate a significant contribution of protected areas (Hoffmann et al. 2010). This could be similarly true for higher plants including CWR. However, in Bolivia there are 22 protected areas established to protect wild populations of flora and fauna, but none consider explicitly CWR in their inventories (SERNAP 2011). According to our study, only one third of the wild potato species endemic to Bolivia (seven species) have been observed to occur within the protected areas. This clearly demonstrates the poor coverage of the actual protected area network in Bolivia in protecting wild potato relatives' populations. As a consequence, an inventory should be made in protected areas that we modelled to have high species richness but have not yet been visited for collecting, particularly 'El Palmar' at the border of Chuquisaca and Cochabamba (Figure 4.3), to get a full understanding as to what extent the existing protected area network in Bolivia can contribute to *in situ* conservation of endemic wild potato diversity. Assisted migration to less-threatened areas, e.g. to existing close-by protected areas, may be an option. We are not aware of examples of such measures, but this option may be worthwhile to explore with the national government body responsible for the protected areas. In a few protected areas also high threat levels were observed (Figure 4.5). So even within these conservation areas, species may be threatened by human disturbance. However, national networks of protected areas are the principle measure for *in situ* conservation of biodiversity. But even protected areas can become susceptible to human pressure. This is of great conservation concern. Several parts of one particular protected area 'Tunari', for example, are severely being threatened according to our analysis. This protected area is close-by to urban populations and people exploit the natural resources in this area (Valenzuela and Padilla Suáre 2002).

The remaining species outside protected areas occur mostly also in natural vegetation habitats. In some case they may occur as weeds in agricultural fields or on the edges of roads, dispersed by human activities. On-farm conservation may therefore be an alternative way to conserve these species, especially those that grow in disturbed areas. Recently, the UNEP/GEF-supported project '*In situ* conservation of wild crop relatives through enhanced information management and field application' (VMABCC-BIOVERSITY 2009) worked on raising awareness of indigenous communities and farmers on the importance of building a participatory conservation strategy for CWR. Guidelines or protocols help raise consciousness and guide farmers in the conservation of CWR (Dulloo et al. 2010). However, there is an on-going discussion about the feasibility to protect CWR on farm, especially how farmers will benefit from this when these wild relatives may not have direct use (e.g. it is the case of strict use of wild potatoes in breeding programs), or even may have negative effects on the productivity of their crops through cross-pollination.

Threat assessment is an important step in setting conservation priorities. In this study, we did that based on threat maps developed by Jarvis et al. (2010). These maps are made on a continental scale and may lose their precision at a local scale. Therefore, these threat analyses are exploratory and where relevant, such as in the area of highest threat levels, a locally more-detailed threat analysis should be carried out. In addition to the observed immediate threats, i.e. accessibility and fire, field observations denote livestock activities pressure as an important threat. This threat has been identified in our analysis as a third immediate threat after accessibility and fire.

Since fire seems to be the most important threat for half of the endemic wild potato species, it would be interesting to investigate how tolerant these species actually are to fire events. Many plant species have adapted to such conditions (Pekin et al. 2009; Ansley et al. 2010; Segarra-Moragues and Ojeda 2010). For them fire may not be a threat and even favour colonization and regeneration. Hijmans et al. (2002) mention that wild potatoes are fire-tolerant. However no further details are provided. It could be that these species can survive fire events underground due to their tubers and resprout in more favourable environmental conditions. On the other hand, human-induced fire events can become so high in frequency and intensity that even ecosystems adapted to natural fire events degrade and thus also the species that inhabit these ecosystems. Ecological research is required to better understand the impact of fire on natural wild potato species. As mentioned above, most collection sites are located in areas of natural vegetation. A possible reason could be that these species do not thrive well in areas disturbed by agriculture. However (Hijmans) et al. 20002 mention that wild potatoes can grow well in disturbed areas albeit they do not explain this in further detail. Another possibility is that there has been a sampling bias towards collecting wild potato species in natural vegetation. It is therefore worthwhile to monitor or set up experiments how well these species may survive in disturbed habitats following conversion to agriculture, which would be relevant for on-farm conservation.

Livestock activities pressure can lead to overgrazing and degradation or replacement of existing vegetation for pastures. Moreover under current climate conditions most ecosystems in which the wild potato species occur are particularly vulnerable to degradation under increased human pressures (Hirota et al. 2011). This probably also has a negative impact on the wild potato populations that inhabit these ecosystems. It can be anticipated that these ecosystems and wild potato populations occurring in these ecosystems will become even more vulnerable under progressive climate change.

This study has identified eight areas where the 21 species could be conserved *in situ*, although this analysis does not take in account the conservation of genetic diversity within species. Endemic species, such as the wild potato species in our study, have in general low levels of genetic diversity within the species whereas relatively high levels of genetic differentiation between their populations can be observed compared to plant species with broader distribution ranges (Hamrick and Godt 1996). Populations of endemic species are therefore susceptible to inbreeding effects and it is important to maintain minimum viable populations. Consequently, the viability of endemic and narrow-distributed species populations may be more sensitive to fragmentation and habitat reduction compared to more wide-spread species. We therefore recommend to carry out population genetic studies on these wild potato species.

At the same time, more wide-spread species may consist of several ecotypes that are adapted to different environmental conditions across the species distribution range. In that case, species populations from the different ecotypes should be conserved to capture this genetic variation. In our study, we found that most of the wild endemic potato species occur in different climate zones. It can be anticipated that these species populations differ in adaptive genetic variation as a response to the local environment under which they have been evolving. Studies of other wild potato species report wide variations in disease resistance between accessions collected in different localities (Ronning et al. 2000; Del Rio et al. 2001). In addition to the prioritized area for species conservation, we identified an additional area in eastern Potosí with high unique ecotype diversity that is relevant to consider in a wild potato conservation strategy.

We followed the classification of Spooner and Salas (2006), which is widely accepted and used in genebanks. However, new taxonomic studies suggest a reduction in the number of species (http://www.solanaceaesource.org). The results of our study would differ substantially if this new taxonomy was followed. Fewer areas would be required to conserve all species. As long as this taxonomic classification is not clarified, we follow the accepted wild potato taxonomy of Spooner and Salas (2006). Molecular characterization studies can help to delineate species and estimate their phylogenetic relationships (González-Orozco et al. 2012). This information provides additional information about genetic distinctiveness to prioritize species for conservation (Weitzman 1998).

4.5 Final remarks

Considering the wide distribution of wild potato species in Bolivia and the often limited resources for germplasm conservation, this study provides guidelines to direct *in situ* conservation efforts to priority areas where there is a higher concentration of species and who have a relatively low level of threat. We prioritized eight areas of about 50 km^2 for species *in situ* conservation, but only one is situated in a protected area, i.e. Area Natural de Manejo Integrado de Apolobamba, where three species are known to occur. A high number of wild potato species is predicted to occur in the protected area 'El Palmar' in north Chuquisaca (Figure 4.3). A field inventory should be carried in that area to assess how many wild potato species it contains.

Ex situ conservation of Bolivian wild potato species is widely represented in 10 genebanks in different countries. Of the 21 endemic species, three are poorly represented in these genebanks, whereas there are no living specimens of two additional species. The protected area 'Area Natural de Manejo Integrado de Apolobamba' has highest priority for additional collecting trips because three of these five species occur in this park. Other areas for targeting collection include La Paz (Province Inquisivi), Cochabamba (Province Ayopaya), the Cochabamba-Santa Cruz and Potosi-Chuquisaca border areas (Figure 4.7).

5. Diversity, distribution, germplasm conservation and use of peach palm (*Bactris gasipaes* Kunth)[i]

Maarten van Zonneveld[1,2]

[1]Ghent University, Faculty of Bioscience Engineering, Coupure links 653, 9000 Gent, Belgium
[2]Bioversity International, Regional Office for the Americas, PO Box 6713, Cali, Colombia

Abstract

Peach palm (*Bactris gasipaes* Kunth) is a multi-purpose palm tree that produces starchy, edible fruits and palm hearts. It may be considered the most important domesticated palm species of the Neotropics and has been widely used since pre-Columbian times. Wild and cultivated peach palm populations are genetically diverse and could offer useful traits for breeding. Changes in land use and climate change pose a serious threat to wild populations *in situ.* While several large *ex situ* field collections of cultivated peach palm accessions exist, these are increasingly difficult to maintain because of the high costs. Screening peach palm diversity for biochemical and morphological traits of commercial and nutritional value would provide a basis for rationalizing collections and enhance future use of peach palm genetic resources. Indeed, well-chosen elite material could then be used either directly for production or in breeding to develop improved peach palm varieties. At the same time, better propagation techniques should be developed to ensure wide distribution of elite peach palm clonal material.

5.1 Introduction

Peach palm (*Bactris gasipaes* Kunth) may be considered the most important domesticated palm species of the Neotropics (Clement et al. 2010). It is a multi-purpose tree that produces starchy, edible fruits and palm hearts. Fruits are rich in starch and contribute a lot to food security and cash income of small-scale farmers who cultivate them. Cultivation and use of peach palm in tropical Latin America has been very popular since pre-Columbian times. This is illustrated by the fact that more than 300 different

---

[i] This chapter has been published in an adapted form as part of the review: Graefe S, Dufour D, van Zonneveld M, Rodriguez F, Gonzalez A (2013) Peach palm (*Bactris gasipaes*) in tropical Latin America: implications for biodiversity conservation, natural resource management and human nutrition. Biodiversity and Conservation 22: 269-300. http://link.springer.com/article/10.1007%2Fs10531-012-0402-3

indigenous names for peach palm have been recorded by chroniclers since the European invasion (Patiño 2002). There are no exact numbers about the importance of peach palm for rural communities, but several local studies have reported the importance of peach palm for subsistence and commercial livelihoods (Graefe et al. 2013). In the northern Peruvian Amazon, for example, at the end of the 20th century, more than 80 % of the producers cultivated peach palm (Labarta and Weber 1998).

The objectives of this chapter are to identify geographic patterns of this species' diversity on the basis of a literature review and to provide recommendations on genetic resources conservation of this tree crop. Several diversity studies have been carried out to better understand the geographic genetic structure and dynamics in use of genetic resources in different parts of the peach palm distribution range. However, at regional level, the geographic patterns of genetic diversity are unclear and its domestication history is open to debate. Such knowledge would help to provide recommendations for peach palm genetic resources conservation on-farm and in wild populations. Changes in land use and climate probably pose a serious threat to wild populations. While several large *ex situ* field collections of cultivated peach palm accessions exist, these are more and more difficult to maintain because of the high costs.

5.2 Botany, reproduction, distribution and domestication

Mapping of georeferenced genebank and herbarium registers obtained from herbaria and genebanks that made their data available through the Global Biodiversity Information Facility (GBIF 2012) and the Brazilian Distributed Information System for Biological Collections (Species Link 2012) showed that cultivated peach palm is currently distributed from Honduras southwards to Central Bolivia and eastwards to Para in Brazil (Figure 5.1). It is usually grown on deep and well-drained soils in areas below 800 masl, with an annual precipitation of 2000-5000 mm and an annual mean temperature above 24° C (Mora-Urpí et al. 1997). Peach palm may occasionally also be found at higher altitudes (up to 1800 masl), as it is the case for the Colombian Cauca region (El Tambo).

Peach palm can be subdivided into a cultivated variety, *Bactris gasipaes* Kunth var. *gasipaes*, and the wild form *Bactris gasipaes* Kunth var. *chichagui* (H. Karsten) Henderson (Henderson 2000). The main difference between the cultivated and wild type is the fruit size (Figure 5.2).

Peach palm is an outcrossing species with unisexual flowers (Mora-Urpí et al. 1997). However, functionally hermaphrodite flowers seem to be occasionally present which would allow selfing under particular conditions (Mora-Urpí et al. 1997). Pollination is mainly through insects, particularly by small curculionid beetles over distances between 100 and 500 meters. However wind and gravity can also function as pollen vector (Mora-

Figure 5.1 Observed peach palm distribution based on herbaria and genebank data made available through GBIF (www.gbif.org) and Specieslink (splink.cria.org.br).

Urpí et al. 1997). There are no clear studies on seed dispersal of wild peach palm. Probably, the latter is restricted to local dispersal by birds and seed-gathering mammals albeit occasionally seed could be dispersed by water, which potentially can lead to seed dispersion over larger geographic distances (Mora Urpí et al. 1997; Clement et al. 2009). Cultivated peach palm fruits seem to be a desired food source for mammals such as rats, squirrels and agoutis that damage many fruits of trees in genebanks (Solano pers. comm., genebank curator of the CATIE peach palm collection; Figure 5.3).

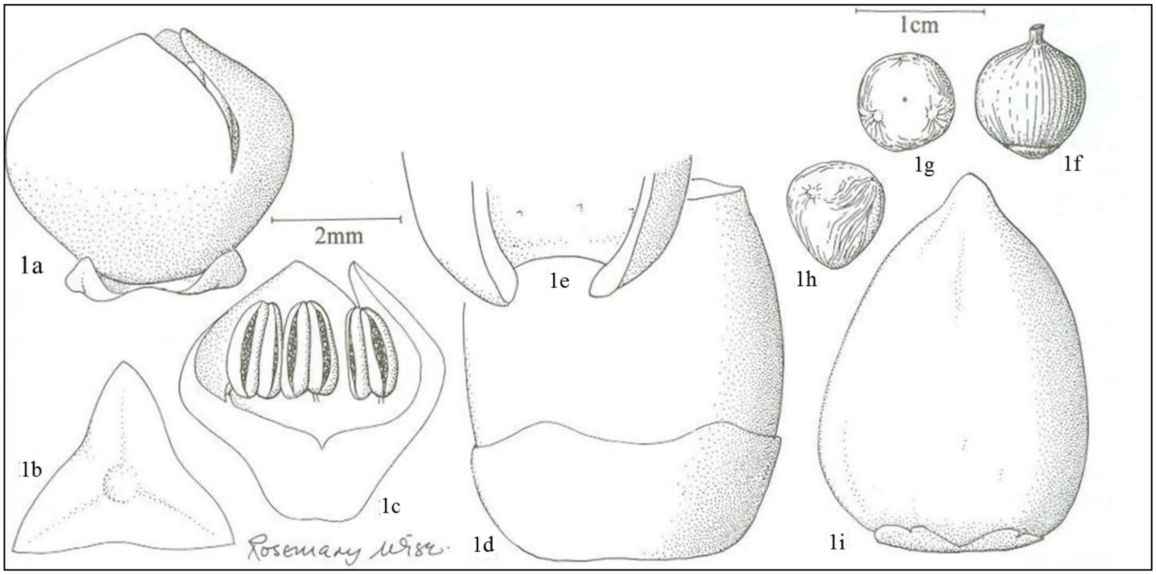


Figure 5.2 1a-1d and 1i: fruit and flower characteristics of *B. gasipaes* var. *gasipaes*; 1f-1h: fruit characteristics of *B. gasipaes* var. *chichagui*; 1a: staminate flower; 1b: sepals; 1c: section of staminate flower; 1d: pistillate flower; 1e: corolla (interior view); 1f: fruit; 1g: endocarp, top view; 1h: endocarp, side view; 1i: fruit (from Henderson 2000).

Phylogenetic studies comparing chloroplast and nuclear DNA polymorphism of species from the *Bactris* clade confirmed a close relationship between cultivated and wild peach palm accessions (Couvreur et al. 2007). Peach palm distribution can be broadly divided based on phenotypic and genetic diversity in (1) two western populations: (1a) Central America, Colombian inter-Andean valleys and Pacific lowlands in Colombia and Ecuador; and (1b) inter-Andean valleys in Maracaibo, Venezuela; (2) and two eastern populations; (2a) the upper Amazon; and (2b) the eastern Amazon (Mora-Urpí et al. 1997; Rodrigues et al. 2004; Hernández-Ugalde et al. 2008).

In general, landraces from the western group are observed to have harder stems, more abundant and stronger spines, larger leaves and more solid rooting in their juvenile phase (Mora-Urpí et al. 1997). The wild form can be subdivided into three types based on differences in fruit types: type I of the southern Amazon; type II of north-eastern Colombia and north-western Venezuela; and type III of the Tropical Andes, south-western Amazon and Central America (Henderson 2000; Clement et al. 2010). Following Henderson (2000), the first type has ‘subglobose fruits with an ostrum and ellipsoid endocarps’. These fruits and endocarps are similar in shape to the bigger fruits of cultivated peach palm (Henderson 2000). Compared to the first type, Henderson describes a second type that has ‘larger, almost ovoid fruits and ellipsoid fruits’ (Henderson 2000). A third type is distinguished that has even smaller fruits and endocarps than the first type (Figure 5.2 1f-g-h; Henderson 2000).

Figure 5.3 Agouti (in captivity) nibbling on a cultivated *Bactris gasipaes* fruit, in Iquitos, Peruvian Amazon.

The exact origin of the cultivated peach palm still remains unresolved. However three hypotheses have been proposed on the origins of its domestication (Clement et al. 2010): (1) a single domestication event in the south-western Amazon, which is supported by phylogenetic studies (Ferreira, 1999) and RAPD marker-data based studies (Rodrigues et al. 2004); (2) a single domestication event in the Colombian inter-Andean valleys and adjacent Pacific lowlands supported by archaeological evidence (Morcote-Rios and Bernal 2001); and (3) multiple independent centres of domestication (Mora Urpí 1999; Hernández-Ugalde et al. 2011).

5.3 Diversity

Considering that peach palm is a long-lived perennial and mainly outcrossing, it can be anticipated that populations and landraces of this species present contain high genetic diversity (Hamrick and Godt 1996; Mora-Urpí et al. 1997). In addition, gene flow and low differentiation is stimulated through extensive documented dispersal routes that may go to a distance of 600 km (Cole et al. 2007). A review of past studies on genetic variation within and between populations using different types of markers and considering allelic richness (A), expected heterozygosity (He) and genetic differentiation (Gst), indeed supports these observations (Table 5.1). No clear areas of high diversity could be identified. Moreover, the fact that different sampling methods, molecular marker techniques and genetic parameters were applied, makes comparison difficult. The use of standardized sets of molecular markers and genetic parameters would certainly help to better understand patterns of genetic variation across peach palm distributions, and find the areas of its domestication (Clement et al. 2010).

Diversity studies confirm the close genetic relationship between wild and cultivated peach palm populations that were identified by Couvreur et al. (2007) in their phylogenetic study. Several studies documented higher similarity between cultivated populations and nearby natural populations than between geographically more distant cultivated populations (Rodrígues et al. 2004; Couvreur et al. 2006; Hérnandez-Ugalde et al. 2008; Araújo et al. 2010). In some cases, however, clear differences could be observed between cultivated populations and wild populations that were used as outlier populations for reference (Silva 2004). One explanation of this close relationship could be that peach palm was domesticated in different locations. With that, the cultivated populations are nowadays still closely related to the nearby natural populations, supporting the hypothesis of multiple origins of domestication (Mora Urpí 1999; Hernández-Ugalde et al. 2011). High similarity between cultivated and natural populations might also be due to introgression between these populations when the domesticated material was introduced into a particular area (Couvreur et al. 2006). Another explanation could be that some of these natural populations are in reality feral populations, i.e. material from cultivated populations that went wild. This has been reported for several old world fruit tree species such as olives, grape, date and figs (Zohary and Spiegel-Roy 1975; Gepts 2003).

The fact that wild and cultivated populations are so closely related suggests that many cultivated peach palm populations are at a semi-domesticated stage. At this stage, introgression with natural populations is still common. While genetic diversity is thus reduced, phenotypic diversity may be enhanced (Clement et al. 2010). Indeed, much phenotypic variation can be observed between and within different cultivated populations (Figure 5.4; Mora-Urpí et al. 1997). Particularly in the upper Amazon, many landraces have been distinguished based on morphological variation whereas they were consequently validated by molecular markers (Sousa et al. 2001; Rodrigues et al. 2004; Silva 2004; Clement et al. 2010). Traditionally cultivated populations can be distinguished in landraces that have (1) fruits lighter than 20 g called ‘microcarpas’ occurring in the eastern and Bolivian Amazon and in the pacific coast of Costa Rica; (2) intermediate fruits called ‘mesocarpas’ that weigh between 20 and 70 g occurring across the whole distribution range; and (3) large fruits titled ‘macrocarpas’ that weigh between 70 and 250 g occurring in the north-western Amazon (Mora-Urpí et al. 1997; Rodrigues et al. 2004; Silva 2004). Fruit size also indicates to which extent a population has been

Table 5.1 Peach palm studies on genetic variation between populations using molecular markers.

Author	Marker type	Nr. of loci	Nr of pop.	Mean nr. individ. per pop.	Type of pop	Covered countries	Mean A per pop.	Highest A	Mean Hes per pop.	highest Hes	Gst
Alves-Pereira et al. (2012)	SSR	11	5	38.4	cultiv.	PE, BR	10.02	Pampa Hermosa, PE (13.10)	0.81	Paranapura, PE (0.83)	0.005
Hérnandez Ugalde et al. (2011)	SSR	5	11	20.50	mixed	BO, BR, CO, CR, EC, PA, PE, VE	6.81	Wild population in Azuero, PA (8.8)	-	-	-
Reis (2009)	SSR	17	11	15.7	mixed	BR, CO, EC, CR, PE, VE	6.86	Cultivated trees from Putumayo, BR/PE (10.82)	0.78	Cultivated trees from Putumayo, BR/PE; Pampa Hermosa, PE; Alto Madeira, BR (0.83)	0.13
Hérnandez Ugalde et al. (2008)	SSR	4	13	38.77	mixed	BO, BR, CO, CR, EC, PA, PE, VE	6.58	Wild population in Azuero, PA (8.75)	0.75	Wild population in Azuero, PA (0.84)	0.15
Cole et al. (2007)	SSR	3	4	55.25	cultiv.	PE	11	San Carlos, PE (12)	0.83	Nuevo San Juan (0.85)	0.001
	SSR	3	4	41.25	cultiv.	PE	11.58	Pucaurquillo, PE (15)	0.79	Puerto Isango (0.83)	0.014
	SSR	3	5	7.4	cultiv.	CO, EC, PE	5.93	Tigre, PE (8.33)	0.76	Putumayo, PE (0.87)	0.003
Couvreur et al. (2006)	SSR	8	4	20.75	mixed	EC, PE, CA	9.23	Cultivated trees from PE and CA (10.70)	0.77	Wild population in EC and cultivated trees from PE and CA (0.80)	
Adin et al. (2004)	AFLP	203	24	10	cultiv.	BR, PE	-	-	0.23	Cultivated trees from San Gabriel de Varadero, PE (0.27)	0.20
Santos et al. (2011)	RAPD	99	6	29	mixed	BR, PE	-	-	0.29	Cultivated trees from Manaus, PE (0.32)	-
Silva (2004)	RAPD	124	10	20	mixed	BR, CO, CR, PA, PE,	-	-	0.25	Cultivated trees from Pará, BR (0.31)	-
Rodrigues et al. (2004)	RAPD	113	9	27.78	mixed	BR, CR, PA, PE	-	-	0.24	Cultivated trees from Solimoes, BR (0.30)	0.16

Cultiv. = cultivated; mixed = cultivated and wild populations

BO=Bolivia; BR=Brazil; CA=Central America; CO=Colombia; CR=Costa Rica; EC=Ecuador; PA=Panama; PE=Peru; VE=Venezuela

A = Average allelic richness per locus; Hes = Average expected heterozygosity per locus; Gst = Genetic differentiation between populations

Figure 5.4 Mature fruit bunches of cultivated peach palm accessions conserved in the peach palm genebank collection of the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) in Costa Rica collected in 2008 (Photo courtesy Scheldeman and Salcedo). They illustrate the wide morphological fruit variation found in cultivated peach palm.

modified following human selection during the domestication process (Clement et al. 2010). Couvreur et al. (2006) identified fruit size as the main characteristic differentiating wild and cultivated peach palm individuals. In a study conducted in Ecuador, they found about 12-33 times bigger fruit volumes of cultivated versus wild individuals (70 vs. 2.1-5.5 cm^3).

Although peach palm is also cultivated in the Guyanas, we could not find information about the existence of particular peach palm landraces or wild populations in this region. Wild Brazilian populations were explored close to the border with French Guiana but these expeditions were unsuccessful (Clement et al. 2009). There is also no evidence available whether this part of the distribution range belongs to an existing population or forms a distinct population.

5.4 Conservation and use of genetic resources

High levels of peach palm phenotypic variation are maintained in *ex situ* germplasm collections, where plant material collected from different areas grow in the same field circumstances (Figure 5.4). Mora Urpí et al. estimated in 1997 a total of 3,309 peach palm accessions with passport data are currently being conserved in 17 collections distributed over eight countries (i.e. Brazil, Colombia, Costa Rica, Ecuador, Nicaragua, Panama, Peru, and Venezuela). A more recent review of only Amazonian peach palm collections reported 2,006 accessions conserved in 10 collections, including also a collection in Bolivia of 200 accessions (Scheldeman et al. 2006).

The maintenance of *ex situ* collections is costly (Clement et al. 2001; Van Leeuwen et al. 2005). Clement et al. (2004) stated that the maintenance of so many and often large collections is not justifiable for an underutilized tree crop such as peach palm. As a consequence, it has been suggested to establish smaller genebanks that better address farmers' needs and consumer preferences (Clement et al. 2004; Van Leeuwen et al. 2005). The establishment of smaller collections that capture most of the genetic variation of the existing germplasm collections could be a good option to reduce maintenance costs (Clement et al. 2001). To assure that these collections are representative for the existing diversity, it is important to screen the latter accessions with molecular markers and on morphological and biochemical characteristics of interest that have high rates of heritability. This is already being done for the peach palm collection of the Instituto Nacional de Pesquisas da Amazônia (INPA) in Brazil (Reis 2009; Araújo et al. 2010).

Most peach palm collections of the Amazon have been characterized either morphologically, molecularly or for both types (Table 5.2, Scheldeman et al. 2006). Several collections have an explicit focus on the characterization of promising material for cooked fruits and flour. The markets for fruit products are above all local peach palm products or only to a lesser extent destined for national or international markets.

Table 5.2 Status of peach palm collections in the Amazon after Scheldeman et al. (2006).

	Germplasm								
	Nr. of accessions	Characterized		Clones selected		Limiting pest and diseases	Agronomic management	Products	Identified markets (local, national, regional, global)
Collection		Yes / No	Objectives	Yes / No	Objectives				
Embrapa - Acre (Brazil)	10	+/-	Identification of promising material	N	-	-	Intermediate	-	Local
Embrapa – Amapá (Brazil)	200	Y	Selection for palm heart	-	-	-	-	-	-
INPA (Brazil)	729	Y	Fruit and palm heart quality	N	-	*Rinchophora* spp.	Intermediate	Palm heart and cooked fruits	Fruits: local; palm heart: national, regional, global
Embrapa – Amazonia Oriental (Brazil)	70 (fruit) 84 (palm heart)	Y	Identification of promising material (morph.)	N	-	-	Intermediate	Palm heart	Fruits: local,; palm heart: national, regional
Embrapa – Roraima (Brazil)	105	+/-	Selection for palm heart	N	-	-	Intermediate	-	Local
Iphae –Bolivia	200	Y	Accessions without spines	+/-	Seed improvement for plants without spines	*Rinchophora* spp. and rodents	Intermediate	Fruit production for cooked fruits, flower, biscuits, liquor and ice-cream	Local
Coorpica - Colombia	50	Y	Identification of promising material	N	-	-	-	-	-
INIAP- Ecuador	121	+/-	Agronomic traits	Y	4 clones for resp. palm heart and fruit quality	-	Advanced (palm heart) Intermediate (fruit)	Palm heart	Fruits: local; palm heart: national, regional, global
INIA/ ICRAF –Peru	350	Y	Production of fruits and resprouts	N	-	*Herminia* spp.	Intermediate	Fruit production for cooked fruits and flower, and palm heart	Local and national
INIA – Venezuela	87	Y	Productivity of all accessions Characterization of 41 accessions (morph. and molec.). Nutritional characterization of 13 accessions	-	-	Termites (Isopteras)	Intermediate	Fruit production for cooked fruits and flower, and palm heart	Local

Characterization is a first step for enhancing the use of accessions that are conserved in genebank collections. Ideally, this should be done in the context of a continuous dialogue between researchers, producers and customers. In this context, participatory domestication of agroforestry species can be a tool to involve small-scale producers. The latter approach allows to ameliorate their livelihoods by sustaining productivity, whilst at the same time conserving on-farm genetic resources (Weber et al. 2001). Within this context, for example, the World Agroforestry Centre (ICRAF) and Peru´s National Institute for Agricultural Research (INIA) initiated in 1997 a participatory genetic improvement program for peach palm heart production and fruit harvesting in the Peruvian Amazon (Weber et al. 2001; Cornelius et al. 2010).

Cultivated populations still contain high levels of diversity in comparison to natural populations and maintain many local human-selected traits (Rodrigues et al. 2004; Couvreur et al. 2006; Hérnandez Ugalde et al. 2008; 2011; Araújo et al. 2010). Low genetic differentiation and exchange of seed material over extensive areas have been documented at least from the Peruvian Amazon (Adin et al. 2004; Cole et al. 2007). Considering also that the rotation length of peach palm as a perennial is large (e.g. more than 10 years), it can be anticipated that there is a low risk of genetic erosion in cultivated populations so that on-farm conservation could be a good alternative for large *ex situ* germplasm collections (Van Leeuwen et al. 2005). However, there is even a risk of genetic erosion for tree crops like peach palm without a proper genetic resources management plan (Cornelius et al. 2006). The same authors compared the effects of different improvement strategies on genetic gain in cultivated peach palm populations and conservation of the species genetic resources in the Peruvian Amazon. Establishment of clonal seed orchards with associated progeny trials based on approximately 450 initial plus trees could be an appropriate strategy to achieve genetic gain while at the same time minimizing genetic erosion. However, this strategy would require vegetative propagation for multiplication (Mora Urpí et al. 1997; Cornelius et al. 2006). Botero Botero and Atehortúa (1999) reported about the development of protocols for the tissue culture technique of somatic embryogenesis in peach palm as a way to maintain genetic pure material. However, there is no indication that this technology is already used for multiplication of selected accessions. This may explain why only in one of the documented Amazonian collections, clones have been selected for propagation (Table 5.2). Nevertheless, research is on-going in order to further improve and automate somatic embryogenesis protocols (Steinmacher et al. 2007; Steinmacher et al. 2011). These protocols are now being tested for the applicability in mass propagation (Steinmacher et al. 2011).

In contrast to cultivated populations, wild peach palm populations, being important resources for genetic improvement, are under threat of deforestation due to agricultural expansion and forest-to-savannah transitions due to climate change in combination with forest degradation (Clement et al. 2009).

The wild peach palm variety does not stand alone; many other Neotropical crop wild relatives are threatened as well (Clement et al. 2009). How these threats affect the three taxonomically different wild peach palm types (see above) is not clear because their distributions are not yet well defined (Clement et al. 2009). Wild peach palm trees are found on disturbed ecosystems, in river banks and primary forest gaps (Mora Urpí et al. 1997). They are often isolated or in low density (Mora Urpí et al. 1997; Da Silva and Clement 2005).

Gene flow of outcrossing tree species with this type of scattered distribution may be restricted. This could result in genetically distinct, isolated subpopulations with small effective population sizes, i.e. the number of reproductive trees in an ideal population under random drift that has the same allelic makeup as the population under study (Mora Urpí et al. 1997). This has implications for defining conservation strategies and requires further research. It would probably be too expensive to conserve *ex situ* a significant amount of wild palm accessions. Instead, it may be more feasible to develop efficient strategies that maximize *in situ* conservation of wild populations. Optimization analyses, such as those proposed by Weitzman (1998), can help to identify which populations can be best conserved *in situ*. Such analyses consider the genetic distinctiveness of each population when compared to other populations and the costs to implement successful conservation measures dependent on the threats of human pressures and progressive climate change. On-farm conservation could be an appropriate alternative for *in situ* conservation of wild populations when high diversity is maintained in nearby cultivated populations and if these cultivated populations are genetically close to wild populations (Hollingsworth et al. 2005). Indeed, this seems to be the case in many parts of the peach palm distribution area (Hérnandez Ugalde et al. 2008). This could also be an adequate, complementary conservation strategy to optimize *in situ* conservation efforts of those wild populations that are genetically most distinct and have a high likelihood to extinct.

To illustrate that PGR conservation in cultivated populations can complement very well *in situ* PGR conservation of natural populations, microsatellite allelic richness of cultivated and natural populations are compared on the basis of microsatellite data (four markers) provided by Hernandez-Ugalde et al. (2008). No differences were found in allelic richness between wild and cultivated populations after comparing ten randomly re-sampled populations of 25 trees without replacement from both the cultivated (n = 220, average 41.2 different alleles per 25 re-sampled trees) and natural gene pool (n = 41, average 42 different alleles per 25 re-sampled trees) (*t* test, $p = 0.4$).

Populations for gene conservation can be prioritized using the reserve selection option in DIVA-GIS that makes use of the Rebelo complementary algorithm (see Scheldeman and van Zonneveld 2010). As already explained in chapter 2 when we were discussing the application of molecular markers in diversity studies, the purpose of this exercise is not to conserve alleles per se. High allelic richness is a proxy for high effective population size with sufficient options for adaptation to environmental changes.

Different alleles and allele combinations may reflect different historical processes of isolation, gene flow and genetic drift (Petit et al. 1998). All alleles (55 in total identified by four markers) can be conserved in a minimum of five populations (Figure 5.5). The highest level of allelic richness (35 alleles) is found in the natural population of Azuero in Panama. Additionally, three cultivated populations are prioritized: Putumayo, Colombia (8 additional alleles), Tembe, Bolivia (5 additional alleles), Tuira, Costa Rica (1 additional allele), and one other natural population: Chontilla, Ecuador (1 additional allele).

Figure 5.5 Prioritization of populations to conserve genetic variation based on the DIVA-GIS complementary reserve selection function using the microsatellite data of Hérnandez Ugalde et al. (2008).

5.5 Conclusions

Both cultivated and wild peach palm populations are genetically diverse. However, on the basis of our findings from literature review no clear geographic patterns of diversity could be detected. Centres of diversity could probably be identified in a genetic diversity study with a sufficient number of sampled trees across the distribution range and loci covered by molecular markers. In the next chapter of this thesis, we will present a case study illustrating how to identify areas of high molecular diversity taking *Annona cherimola* (Mill.) as a case study. This species has been extensively sampled across its Andean distribution range. Cultivated populations that are genetically closely related to threatened wild populations could have an important complementary role in conservation of these wild genetic resources. *Ex situ* collections of material from cultivated populations can conserve a fair amount of diversity but are costly to maintain. Screening peach palm diversity for biochemical and morphological traits of commercial and nutritional value could provide a basis for rationalizing collections and enhance the use of peach palm genetic resources. Elite material could be used either directly for production or in breeding to develop improved peach palm varieties. Materials showing traits of interest could be conserved *in situ* through the establishment of local clonal or seed orchards. At the same time, better propagation techniques should be developed to ensure wider distribution of elite peach palm clones.

6. Mapping genetic diversity of cherimoya (*Annona cherimola* Mill.): application of spatial analysis for conservation and use of plant genetic resources[i]

Maarten van Zonneveld[1,2], Xavier Scheldeman[2], Pilar Escribano[3], María A. Viruel[3], Patrick Van Damme[1,4,5], Willman Garcia[6], César Tapia[7], José Romero[8], Manuel Sigueñas[8] and José I. Hormaza[9]

[1] Ghent University, Faculty of Bioscience Engineering, Coupure links 653, 9000 Gent, Belgium
[2]Bioversity International, Regional Office for the Americas, PO Box 6713, Cali, Colombia
[3] Instituto de Hortofruticultura Subtropical y Mediterránea, (IHSM-UMA-CSIC), Estación Experimental La Mayora, 29750 Algarrobo-Costa, Málaga, Spain
[4] World Agroforestry Centre (ICRAF), GRP1 - Domestication, PO Box 30677, Nairobi 00100, Kenya
[5]Institute of Tropics and Subtropics, Czech University of Life Sciences, Prague, Kamycka 129, Prague 6, Suchdol, 165 21, Czech Republic
[6] PROINPA, Oficina Regional Valle Norte, Av. Meneces s/n. Km. 4 Zona El Paso, Cochabamba, Bolivia
[7] Instituto Nacional Autónomo de Investigaciones Agropecuarias (INIAP) Panamericana sur km1, Quito, Ecuador
[8] Naturaleza y Cultura Internacional (NCI), Av. Pío Jaramillo A. y Venezuela, Loja, Ecuador
[9] Instituto Nacional de Innovación Agrícola (INIA), Av. La Molina No. 1981 La Molina, Lima, Peru

Abstract

Geographic Information Systems (GIS) can be used to allow visualization and better understanding of spatial patterns of genetic diversity. These would then serve as key input to optimize conservation and use of plant genetic resources. In this chapter, we

---

[i] Adapted from: van Zonneveld M, Scheldeman X, Escribano P, Viruel MA, Van Damme P et al. (2012) Mapping Genetic Diversity of Cherimoya (*Annona cherimola* Mill.): Application of Spatial Analysis for Conservation and Use of Plant Genetic Resources. PLoS ONE 7: e29845. http://dx.plos.org/10.1371/journal.pone.0029845

explore the possibilities of incorporating molecular marker characterization data into GIS. To do this, we develop on a case study of cherimoya (*Annona cherimola* Mill.), a Neotropical fruit tree species. We present spatial analyses to (1) improve the understanding of spatial distribution of genetic diversity of cherimoya natural stands and cultivated trees in Ecuador, Bolivia and Peru identified with microsatellite molecular markers (SSRs); and (2) formulate best conservation strategies by revealing priority areas for *in situ* conservation, and identifying existing diversity gaps in *ex situ* collections. We found high levels of allelic richness, locally common alleles and expected heterozygosity in cherimoya's putative centre of origin, i.e. southern Ecuador and northern Peru, whereas levels of diversity in southern Peru and especially in Bolivia were significantly lower. The application of GIS on a large microsatellite dataset allows for a prioritization of more restricted areas for *in situ* conservation and targeted collection across the Andean distribution range of cherimoya than previous studies could do, i.e. at province and department level in Ecuador and Peru, respectively.

6.1 Introduction

In this chapter, we present a case study on the geographic distribution and genetic diversity of cherimoya in its Andean distribution range in order to test and exemplify the usefulness of combining molecular marker techniques and spatial data to inform *in situ* conservation decisions.

Cherimoya is an underutilized Neotropical fruit tree species belonging to the Annonaceae, a family included within the Magnoliales in the Eumagnoliid clade among the early divergent angiosperms (Bremer et al. 2009). The species is still in initial stages of domestication (Escribano et al. 2007). It is considered at high risk of losing valuable genetic material from its gene pool (National Research Council 1989b). Around Quito, for example, most of the traditional cherimoya cultivation is being replaced by avocado plantations, which are commercially more attractive (Scheldeman, pers. obs.).

Cherimoya fruits are widely praised for their excellent organoleptic characteristics. The species is therefore considered to have high potential for commercial production and income generation for both small and large-scale producers in subtropical climates (Van Damme and Scheldeman 1999). Cherimoya presents protogynous dichogamy, i.e. it has hermaphroditic flowers wherein female parts mature before the male parts, favouring outcrossing in its native range (Lora et al. 2010). For commercial production outside of the tree's native range, hand pollination is a common and needed practice due to lack in overlap of female and male stages, and absence of pollinating agents (Lora et al. 2010). To do this, first pollen of pollen-shedding flowers is collected. This can be conserved for maximum three days (Scheldeman 2002). The pollen is used to fertilize other freshly opened flowers, i.e. the stage at which the female function is mature (Schroeder 1941). At present, large-scale commercial production is mainly concentrated in Spain, the world's largest cherimoya producer, with around 3,000 ha of plantations. On the other hand, small-scale cultivation occurs throughout the Andes, Central America and Mexico.

Cherimoya is grown in a wide climate and altitudinal range. Edaphoclimatic studies performed in cherimoya (Bydekerke et al. 1999; Farré et al. 1999; Van Damme et al. 2000) suggest that cherimoya grow in the subtropical strata (1,300 – 2,300 masl) of the Neotropics characterized by a dry winter and a wet summer; optimum mean annual temperature ranging from 16 ºC to 20 ºC and rainfall between 650 and 1,250 mm per year. Temperatures above 30 ºC usually result in pollination problems and could cause burnings in leaves and fruits and promote the drop of recently set fruits whereas temperatures below -2 ºC can produce damage in leaves, fruits and trunks. The reproductive process is especially sensible to temperature and humidity changes (Lora et al. 2009; 2011; 2012).

Most early chroniclers and scientists proposed the Andean region, and more specifically the valleys of southern Ecuador and northern Peru, as cherimoya's centre of origin (Popenoe 1921; National Research Council 1989b). The occurrence of isolated, putatively wild cherimoya forest patches in the inter-Andean valleys of Ecuador and northern Peru supports this hypothesis. Nonetheless, the possibility that these are feral populations cannot be excluded. Ferality is a wide-spread phenomenon in annual and perennial crops (Ellstrand et al. 2010). It has also been observed for several fruit tree species, such as olives, figs, dates (Zohary and Spiegel-Roy 1975; Gepts 2003).

An alternative hypothesis for the centre of origin of cherimoya is Central America because most relatives of cherimoya are native to Central America and southern Mexico (Rainer, Institute of Botany, University of Vienna, 2011, pers. comm.). In addition, a high genetic diversity is found in cherimoya genotypes from that area (Hormaza et al., unpublished data). In any case, cherimoya fruits have been consumed in the Andean region since antiquity (National Research Council 1989b) and movement of germplasm across southern Mexico, Mesoamerica and the Andes probably took place already in pre-Columbian times. Wolters (1999) advocated that the ceramic cherimoya-shaped vases found at archaeological sites of the Ecuadorian Valdivia culture (5,500 – 3,600 years ago) may testify of the important role this ancient culture played in exchange of cherimoya germplasm and other crops between the Andean region and Mesoamerica.

The conservation status of cherimoya genetic resources has improved considerably in recent years. Due to an increase in commercial prices for cherimoya at local markets, Andean farmers are motivated to conserve *in situ* the cherimoya trees growing in their backyards. Indeed, trees present in home gardens and orchards are common throughout the Andean region in Bolivia, Ecuador and Peru. They usually originated from seeds taken of local material or from chance seedlings (Escribano et al. 2007). Among them, some individuals show promising traits for future breeding programs (Scheldeman et al. 2003). In Peru, the local selection 'Cumbe' is already fetching retail prices significantly above the prices of unselected cherimoya fruit types (Vanhove and Van Damme 2009).

In contrast to most tropical and subtropical underutilized fruit tree species, cherimoya genetic resources are well-conserved *ex situ*. Several field collections have been established in Spain, Peru and Ecuador, comprising over 500 different accessions (Escribano et al. 2007; CHERLA 2008). The Spanish collection, based at la Estación Experimental La Mayora in Malaga, and holding over 300 accessions (190 collected in the Andean region), is currently used as source materials for the Spanish cherimoya breeding program and has been thoroughly analysed using isozymes (Pascual et al. 1993; Perfectti and Pascual 1998; 2005) and microsatellite markers (Escribano et al. 2004; 2008a; 2008b).

The recent development of new molecular tools in combination with new geospatial methods and increased computer capacity has created opportunities for new applications of genetic diversity analyses (Manel et al. 2003; Holderegger et al. 2010; Scheldeman and van Zonneveld 2010). Whereas neutral molecular markers are considered a sound tool to measure patterns and trends in the use and conservation of plant genetic resources (Eaton et al. 2006), Geographic Information Systems (GIS) provide opportunities to carry out spatial analyses of genetic diversity patterns identified with these markers (Kozak et al. 2008). GIS can be used to interpolate genetic parameters between sampled populations (e.g. Degen and Scholz 1998; Hanotte et al. 2002; Hoffmann et al. 2003); to apply re-sampling of georeferenced samples within a defined buffer zone (Lowe et al. 2000; Vigouroux et al. 2008); or to develop grid-based genetic distance models (McRae 2006; van Etten and Hijmans 2010). GIS are also an acknowledged tool to prioritize areas for conservation of plant genetic resources (Guarino et al. 2002). Several studies have used geospatial analysis to develop conservation strategies for plant genetic resources based on molecular marker characterization data (e.g. Lowe et al. 2000; Kiambi et al. 2008). Moreover, results obtained using GIS can be presented in a clear way through maps. This facilitates the incorporation of these findings into the formulation of conservation strategies and the implementation of conservation measures (Jarvis et al. 2010).

The specific objectives of this research are to (1) apply innovative geospatial analysis to improve understanding of the geographic distribution of cherimoya‘s genetic diversity in its putative native range, identified with microsatellite molecular markers (SSRs); and (2) formulate the best-possible conservation strategies by prioritizing areas for conservation both on-farm and in wild populations, and identifying existing diversity gaps in *ex situ* collections. Based on the outcomes, we discuss how these spatial approaches can be used to define strategies that will guarantee sustainable long-term conservation of cherimoya genetic resources and how they can be applied to improve conservation and use of tree and crop genetic resources in general.

6.2 Results

A total of 1,504 trees were analysed in this study, i.e. 395 from Bolivia, 351 from Ecuador and 758 from Peru. Of those, 502 are currently conserved in *ex situ* collections (either in Ecuador, Peru or Spain) whereas the remainder trees were sampled *in situ* between 2006 and 2008. The molecular analysis included a core set of nine microsatellite loci (Escribano et al. 2008a) resulting in 71 different alleles. In all analyses of *alpha* diversity and *beta* diversity (also referred to as divergence) we applied circular neighbourhood re-sampling technique resulting in a total dataset of 48,128 trees (Figure 6.1). This technique facilitates analysis of patterns in genetic variation across extensive distribution ranges while maintaining high-resolution grids. In this result section, we will introduce several technical concepts. We will come back in more detail to the materials and methodology used in this chapter, in section 6.4 after the discussion.

Allelic richness

Allelic richness is a straightforward measure of genetic diversity that is commonly used in diversity studies on the basis of molecular markers that aim at selecting populations for conservation because it is an indicator of effective population size, expressing the rate of historic gene flow and bottleneck events (Frankel et al. 1995a; Petit et al. 1998). Figure 6.2 presents the distribution of the average number of alleles per locus found in the study area. It clearly shows that a higher number of alleles is present in the northern part of the study area, specifically in northern Peru, around Cajamarca Department. Other areas of high diversity are located on the border zone between Ecuador (Loja Province) and Peru (Piura Department), in the northern part of Ecuador around its capital Quito and in the northern part of the Lima Department in Peru.

Allelic richness corrected by rarefaction

Despite the effort to apply a similar sampling density throughout the study area, some areas (often locations with a higher abundance of traditionally managed cherimoya trees and stands) were sampled more intensively than others (Figure 6.1), generating a sampling bias (Hijmans et al. 2000). The rarefaction methodology corrects this sampling bias by recalculating allelic richness in each grid cell to a minimum sample size (Petit et al. 1998). Figure 6.3 shows only grid cells where 20 or more trees were present after applying a one-degree circular neighbourhood approach, and for which allelic richness was corrected following the rarefaction methodology to a minimum sample size of 20 trees. This is similar to the sample size that Petit et al. (1998) used to estimate allelic richness in each population. Cajamarca Department in northern Peru remains the area with the highest diversity, up to an average of 5.18 different alleles per locus.

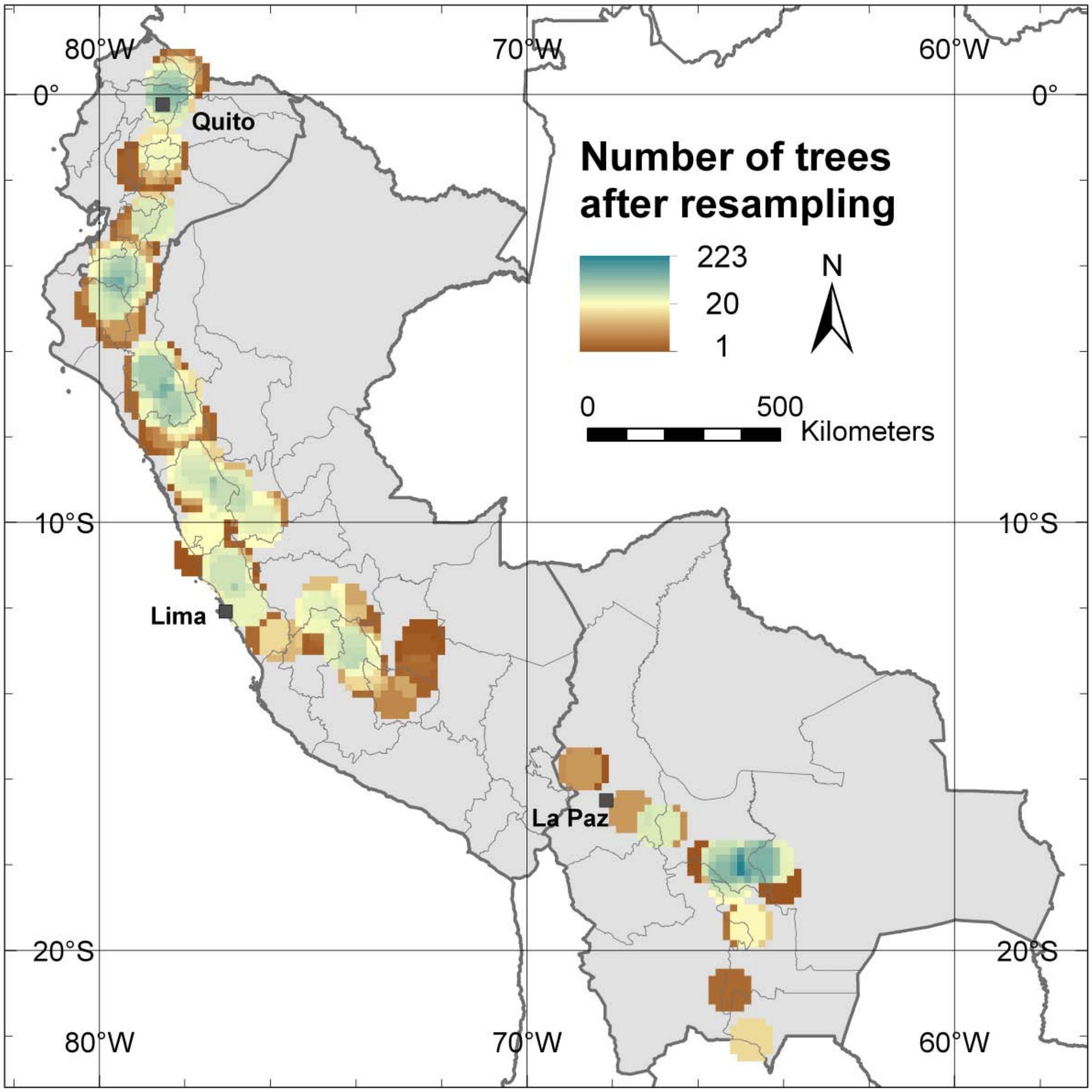


Figure 6.1 Number of cherimoya (*Annona cherimola*) trees per 10-minutes grid cell after re-sampling applying one-degree circular neighbourhood technique.

After correction by rarefaction, diversity in Ecuador, especially around Quito, is reduced. The same seems to happen in the northern part of the Lima Department, in Peru. This evidences the presence of a sampling bias around the capitals of both countries. The area around the Peruvian capital Lima, an important commercial cherimoya cultivation area, shows the lowest allelic richness within Peru. This can probably be explained by the widespread cultivation of a vegetatively propagated cultivar, ‘Cumbe’. Another striking result is that allelic richness in Bolivia, which was already low in the uncorrected analysis, is even lower with correction for sampling bias. This results in an even higher contrast between cherimoya genetic diversity in Bolivia and that found in Peru and Ecuador.

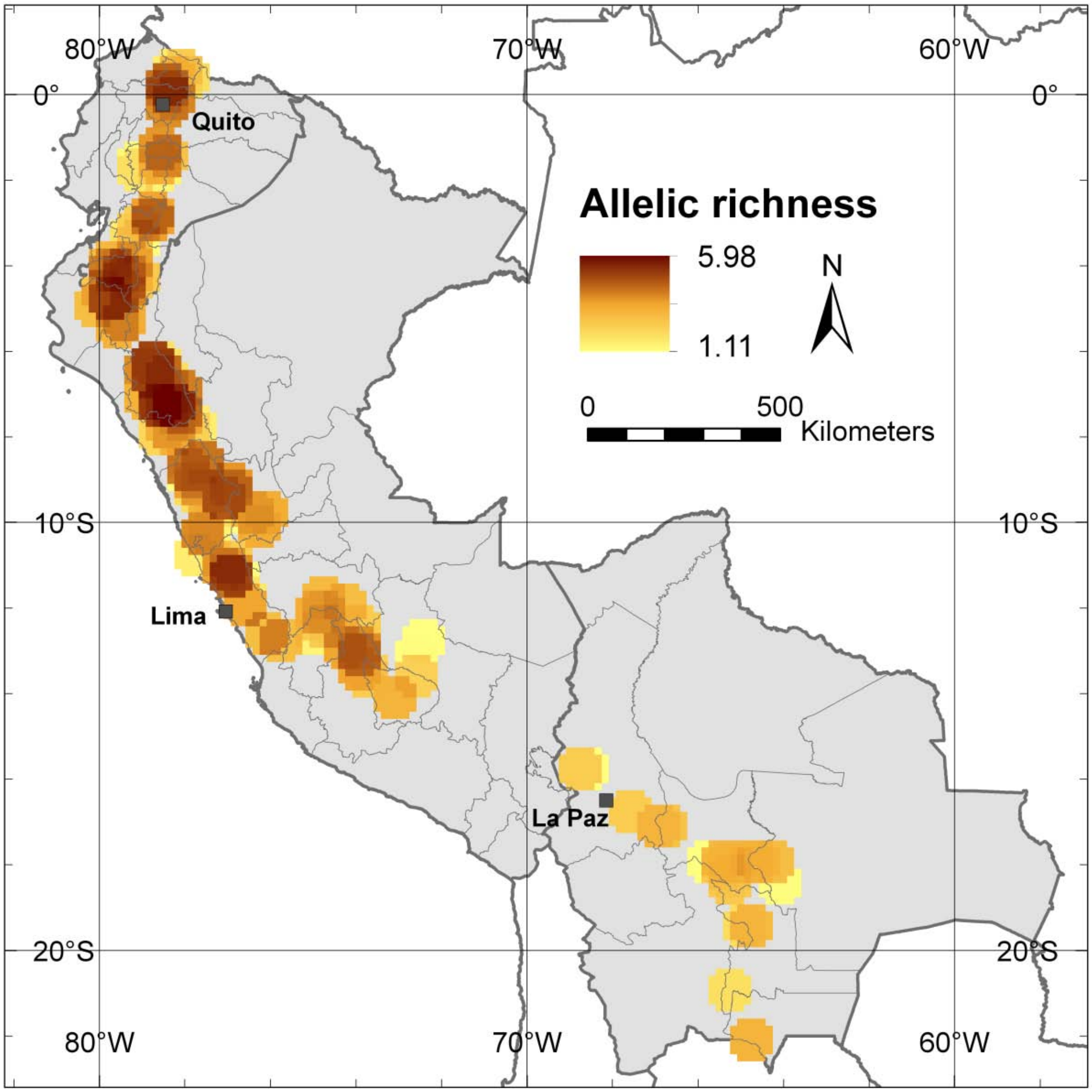


Figure 6.2 Average number of alleles per locus for cherimoya (*Annona cherimola)* trees in all 10-minutes grid cells applying one-degree circular neighbourhood technique.

Locally common alleles

Priority in conservation should be given to populations that contain locally common alleles. These are alleles that occur in high frequency over a limited area, and evidence the presence of genotypes adapted to specific environments and long histories of local natural and human selection (Frankel et al. 1995a; van de Wouw et al. 2010a). Figure 6.4 shows the richness of locally common alleles per locus in the study area. The high diversity levels found in the Cajamarca Department in northern Peru are reconfirmed.

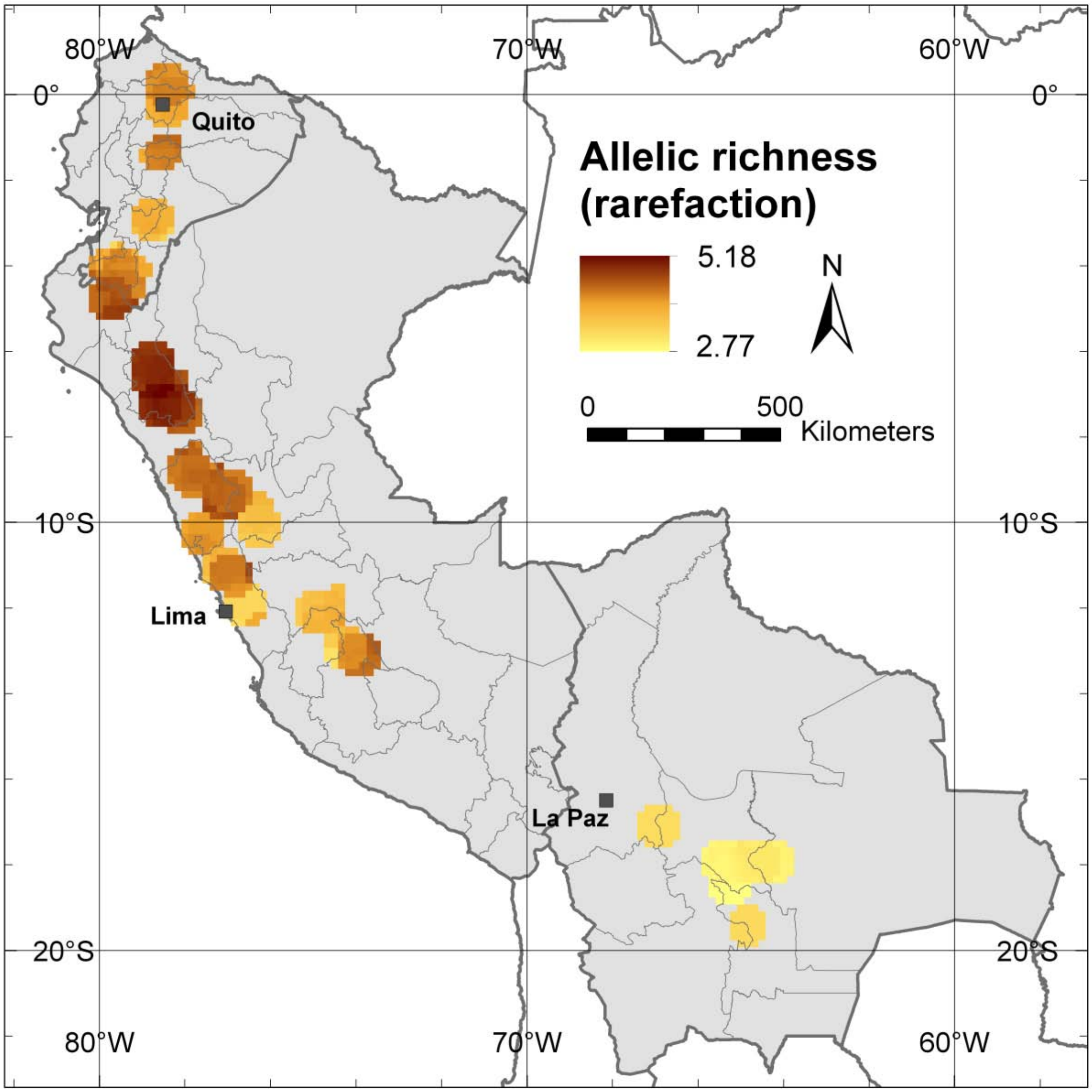


Figure 6.3 Average number of alleles per locus in 10-minutes grid cells applying a one-degree circular neighbourhood technique and a correction by rarefaction to a minimum sample size of 20 trees.

Besides harbouring the highest number of different alleles, this area also contains the highest number of locally common alleles. This makes this area a priority for *in situ* conservation, both of cultivated trees on-farm and in natural stands. The border region between Peru and Ecuador (Piura Department and Loja Province) is another area where a high concentration of locally common alleles has been observed. It may, therefore, be a second area to prioritize *in situ* conservation efforts. To a lesser extent, the area around Quito in Ecuador and the northern part of the Lima Department in Peru also present locally common alleles.

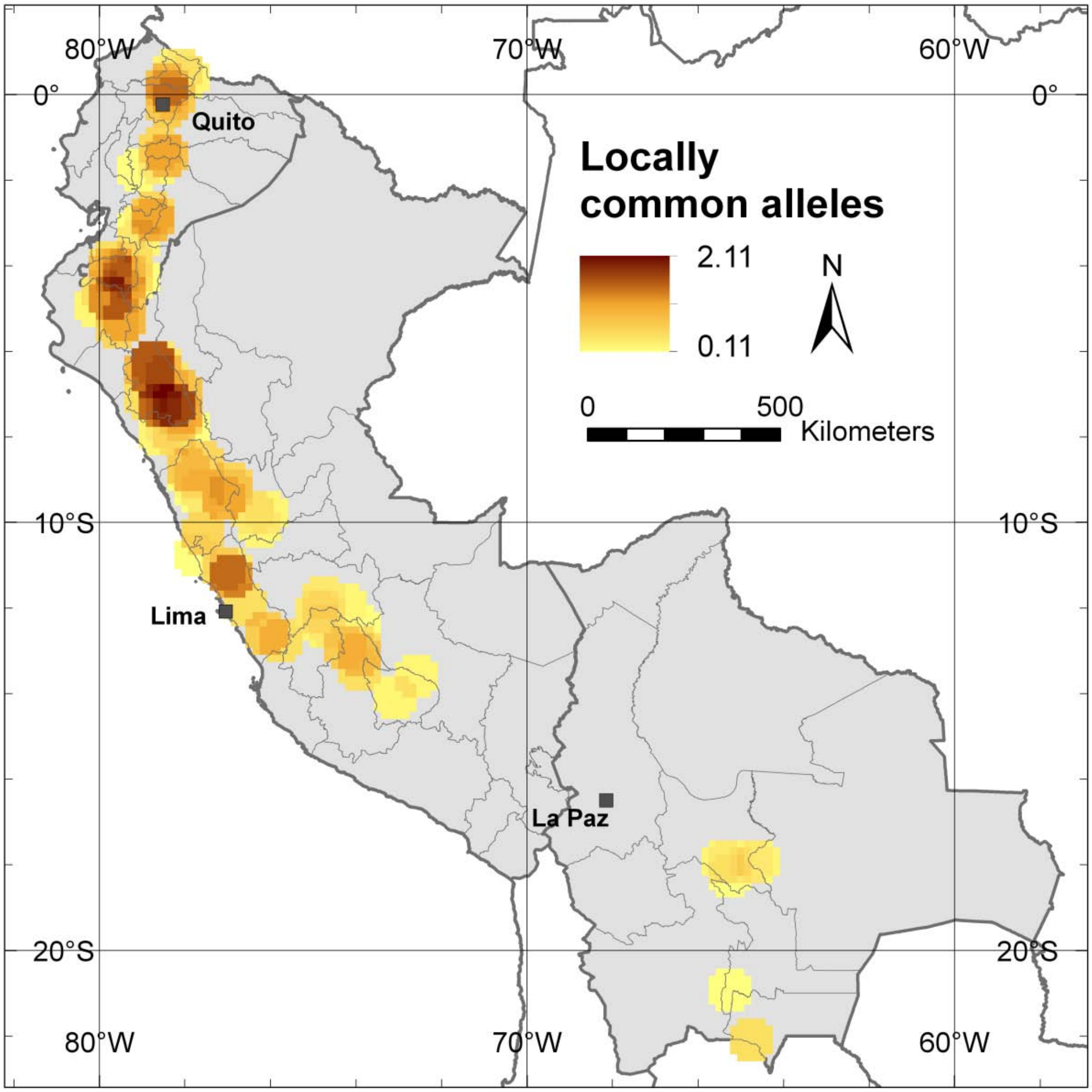


Figure 6.4 Average number of alleles per locus that are relatively common (occurring with a frequency higher that 5 %) in a limited area (in 25 % or less of the grid cells) in 10-minutes grid cells applying one-degree circular neighbourhood technique.

*Expected Heterozygosity (*He*) and Fixation Index (*F*)*

In situ conservation should focus on viable populations, where inbreeding and subsequent loss of alleles are minimal. Parameters that allow assessment of inbreeding are expected heterozygosity (*He*) and the fixation index (*F*) (Peakall and Smouse 2006). The latter parameter helps to detect areas subjected to high inbreeding depression and, as the inverse to that, excess in heterozygosity (Peakall and Smouse 2006).

Figure 6.5 shows the values for *He* in the study area, again confirming Cajamarca Department in northern Peru as the area with the highest genetic diversity. High *He*

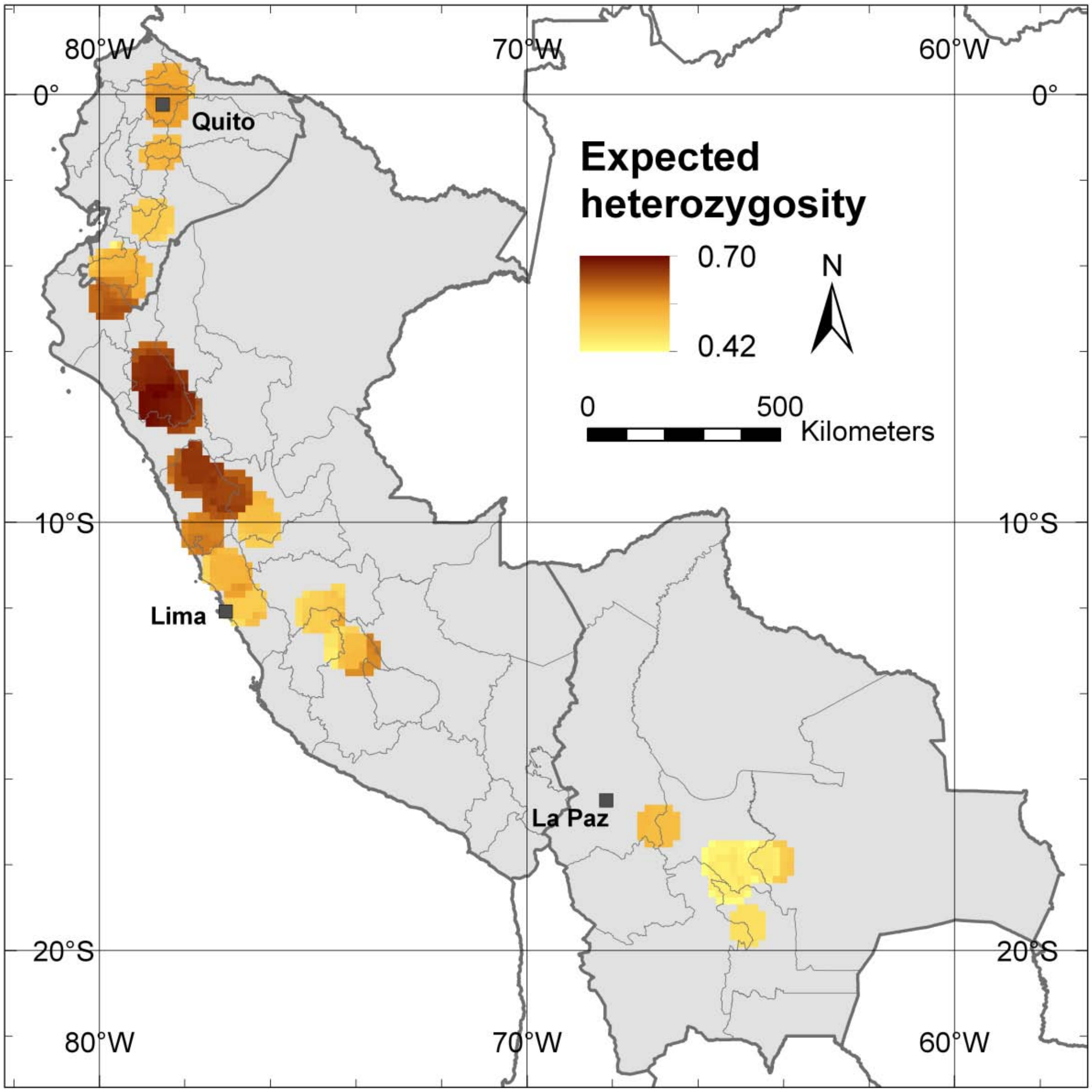


Figure 6.5 Average *He* value for cherimoya (*Annona cherimola*) trees in each 10-minutes grid cell with 20 or more trees applying one-degree circular neighbourhood.

values, however, radiate towards the south (as opposed to the higher diversity towards the north found in the allelic richness analyses) indicating higher levels of diversity in terms of heterozygosity in central Peru compared to Ecuador.

Figure 6.6 shows the values for the fixation index, with *F* values close to 0 in Cajamarca Department indicating that natural and cultivated cherimoya tree stands in this area have not experienced much or any inbreeding. The highest values for *F* are observed in central Ecuador, suggesting that the level of inbreeding is highest in that part of cherimoya's Andean distribution range.

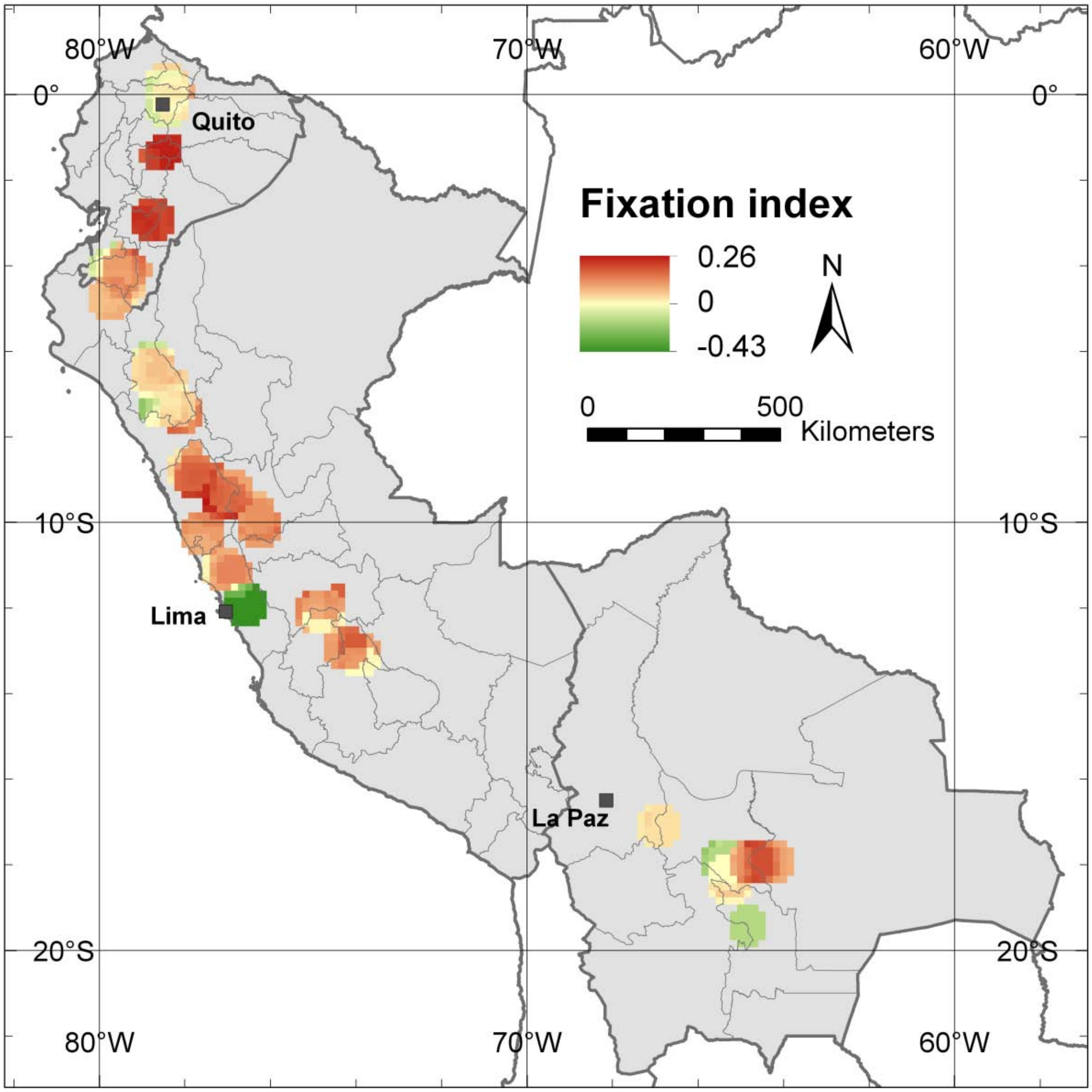


Figure 6.6 Average *F* value in each 10-minutes cell with 20 or more trees applying one-degree circular neighbourhood technique. Yellow areas indicate cherimoya stands where observed heterozygosity is as expected; red areas indicate stands where observed heterozygosity is lower than expected (indicating inbreeding) whereas observed heterozygosity is higher than expected in green areas.

*Genetic Distance (*GD*) to the local cultivar Cumbe*

The most important Peruvian commercial cherimoya cultivation area, located near the capital Lima, particularly shows negative *F* values, i.e. an excess of heterozygosity. Most cherimoyas cultivated in this area are vegetatively propagated clones of cultivar ‘Cumbe’.

This resulted in highly heterozygous values from the molecular analysis, i.e. the ‘Cumbe’ accession conserved in the Spanish genebank is heterozygote for eight of the

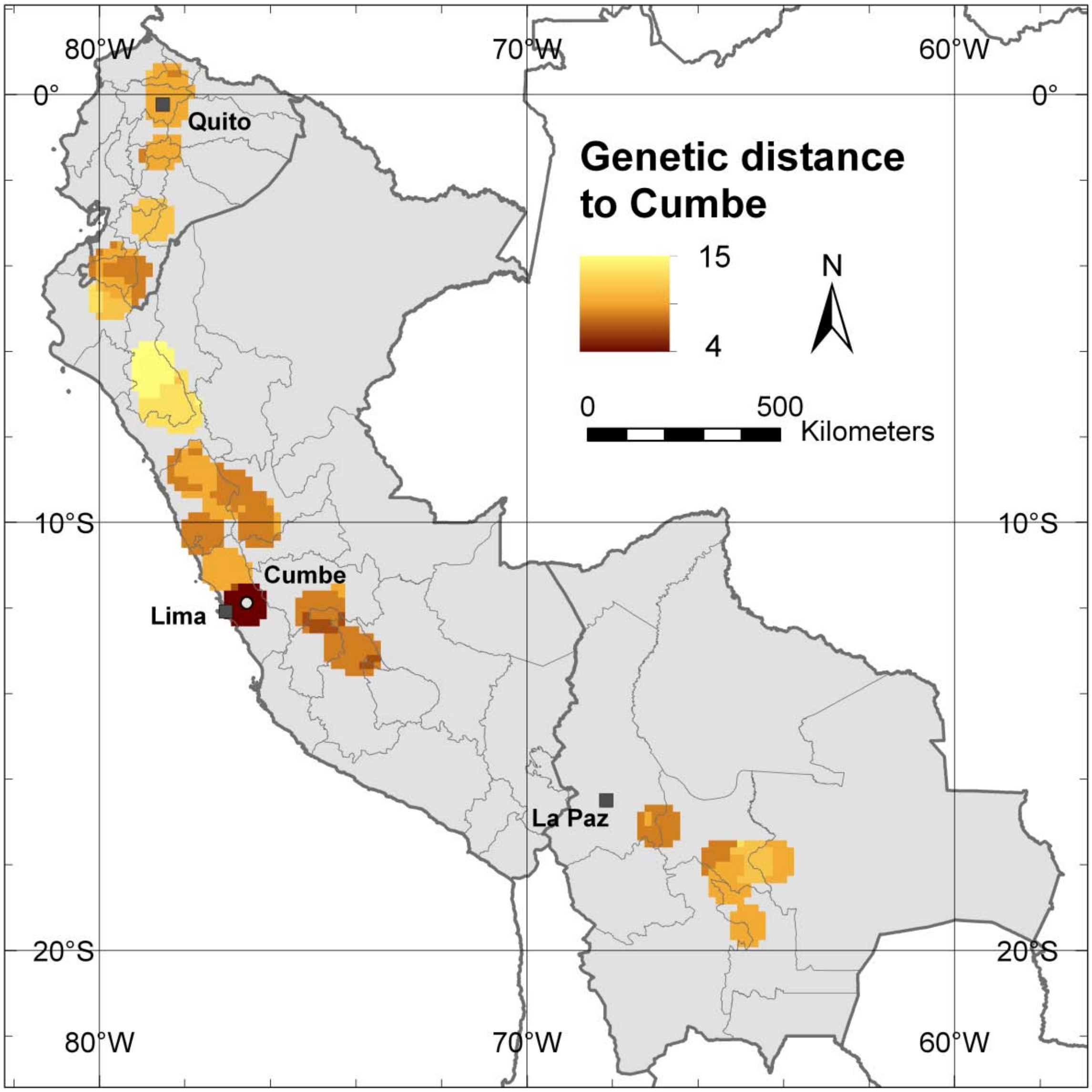


Figure 6.7 Average genetic distance (GD) to cultivar 'Cumbe', in each 10-minutes cell with 20 or more trees, applying one-degree circular neighbourhood technique. The 'Cumbe' accession from the collection la Mayora, Malaga, Spain, was used as a reference of the cultivar.

nine microsatellite loci analysed in this study (*Ho* value of 0.89). An analysis of the average genetic distance, between the 'Cumbe' accession and the genotypes in each grid cell with 20 or more re-sampled trees in the study area, clearly shows lowest genetic distance values near the Peruvian capital, Lima, indicating that the cherimoya trees in this area are very similar to cultivar 'Cumbe' (Figure 6.7). This area clearly differs from the rest of the cherimoya distribution area in our study. The genetic patterns in other areas are likely to be a product of more natural gene flow patterns.

Beta diversity (divergence)

Besides *alpha* diversity parameters, aimed at identifying those areas with highest allelic richness and balanced allele frequencies, *in situ* conservation also needs to take into account allelic composition (*beta* diversity or divergence) as it is possible that populations with low allelic richness possess unique allele compositions that are different from those of populations in other areas of the range, which would warrant their *in situ* conservation (Petit et al. 1998). Breeding possibilities between the ancestors of current trees from different geographic clusters may have been restricted historically. Dependent on the time that has passed since these limitations in mating; their ancestors can have undergone other natural and human selection processes resulting in differences in functional traits. Applying the Structure software (see Pritchard et al. 2000) and using the statistic parameter ΔK following Evanno et al. (2005) to define the number of clusters with genetically similar trees present in the study area, we differentiated two main populations. Figure 6.8 shows the differentiation of the populations among distribution areas in cluster A and B, respectively. Cluster A has the highest presence in the areas previously identified as those with the highest allelic richness (Cajamarca Department in northern Peru; border zone between Ecuador and Peru; and the area around Quito in Ecuador), whereas cluster B is mainly confined to southern Peru and Bolivia. Bolivian cherimoya trees are almost exclusively assigned to cluster B. Particular areas that did not show a strong linkage to either of the two clusters included the surroundings of the city of Lima and Loja Province in southern Ecuador.

Ex situ *conservation status*

Of the 1,504 trees included in this study, 502 genotypes are currently conserved in *ex situ* collections (either in Ecuador, Peru or Spain). Only eight alleles, corresponding to 11 % of the total of 71 alleles that have been found in the study area, are not represented in any accession of these collections. Figure 6.9 shows the distribution of the missing alleles.

There is only a small area with a significant portion of missing alleles (3 in total), i.e. in southern Ecuador (Azuay Province). Natural cherimoya forest patches and areas of traditional cherimoya cultivation in this province should be prioritized for future cherimoya collection missions. With almost 90 % of alleles found to be present in *ex situ* collections, it can be concluded that, in general, cherimoya diversity from the countries analysed is fairly well conserved *ex situ*.

Distribution range of cherimoya in the Andes

The above results and subsequent conclusions are obviously only of practical use if the sampling performed was indeed representative for the distribution of cherimoya in the

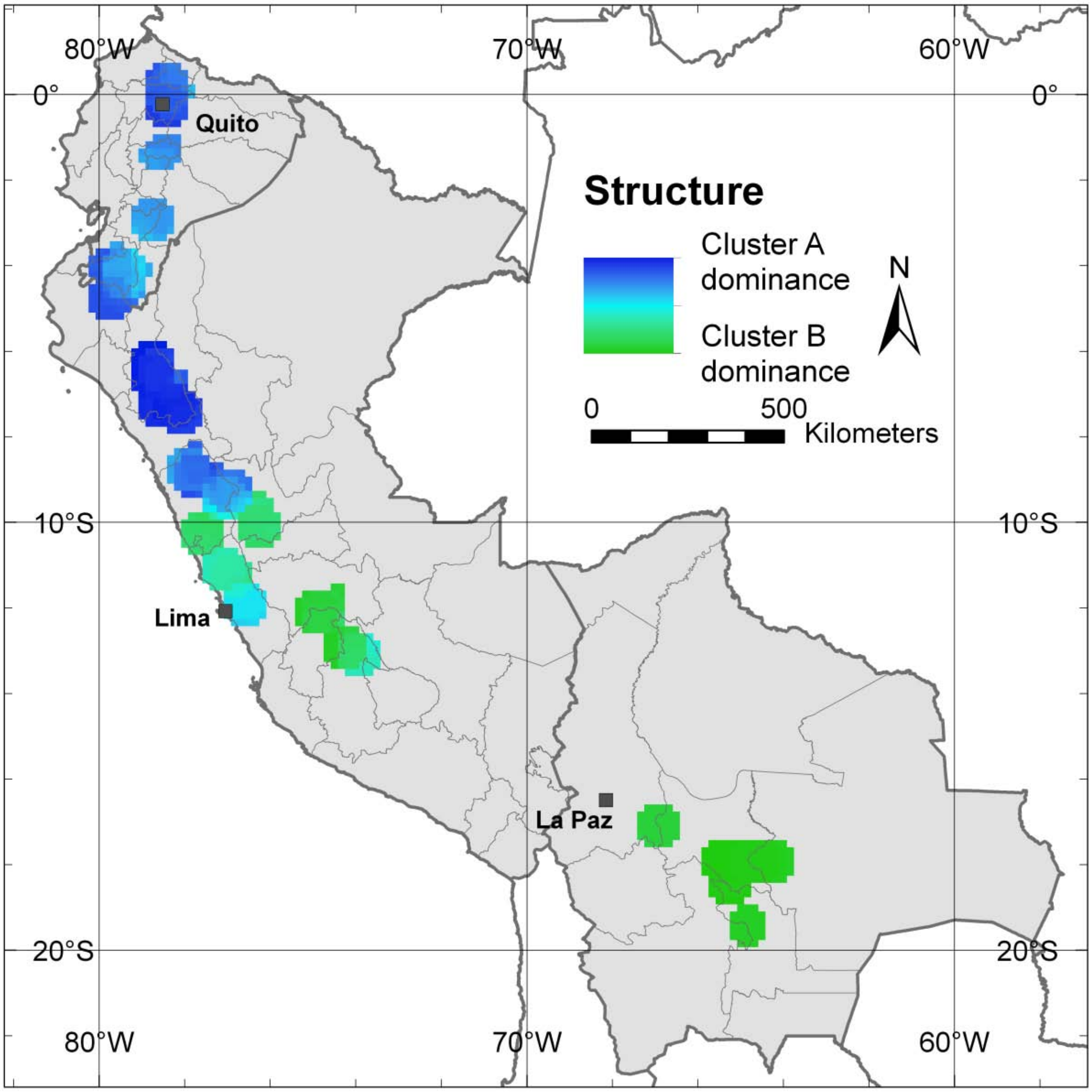


Figure 6.8 Average probability of finding a cherimoya tree belonging to cluster A or B in each 10-minutes cell with 20 or more trees applying a one-degree circular neighbourhood. Dark blue areas show a higher probability of finding trees belonging to cluster A, whereas dark green areas show a higher probability of finding trees belonging to cluster B. Light blue-coloured areas are not clearly assigned to any of the two clusters.

study area. Environmental Envelope Modelling (EEM) with Maxent software was applied to model cherimoya's distribution range in Ecuador, Peru and Bolivia based on the climatic niche in which the 1,504 sampled trees of our study were located. The modelled distribution was then compared to that of the sampled areas in these countries.

Cross-validation, to evaluate the quality of the distribution model, returned an Area Under Curve (AUC) value of 0.9, which indicates good model performance (Araújo

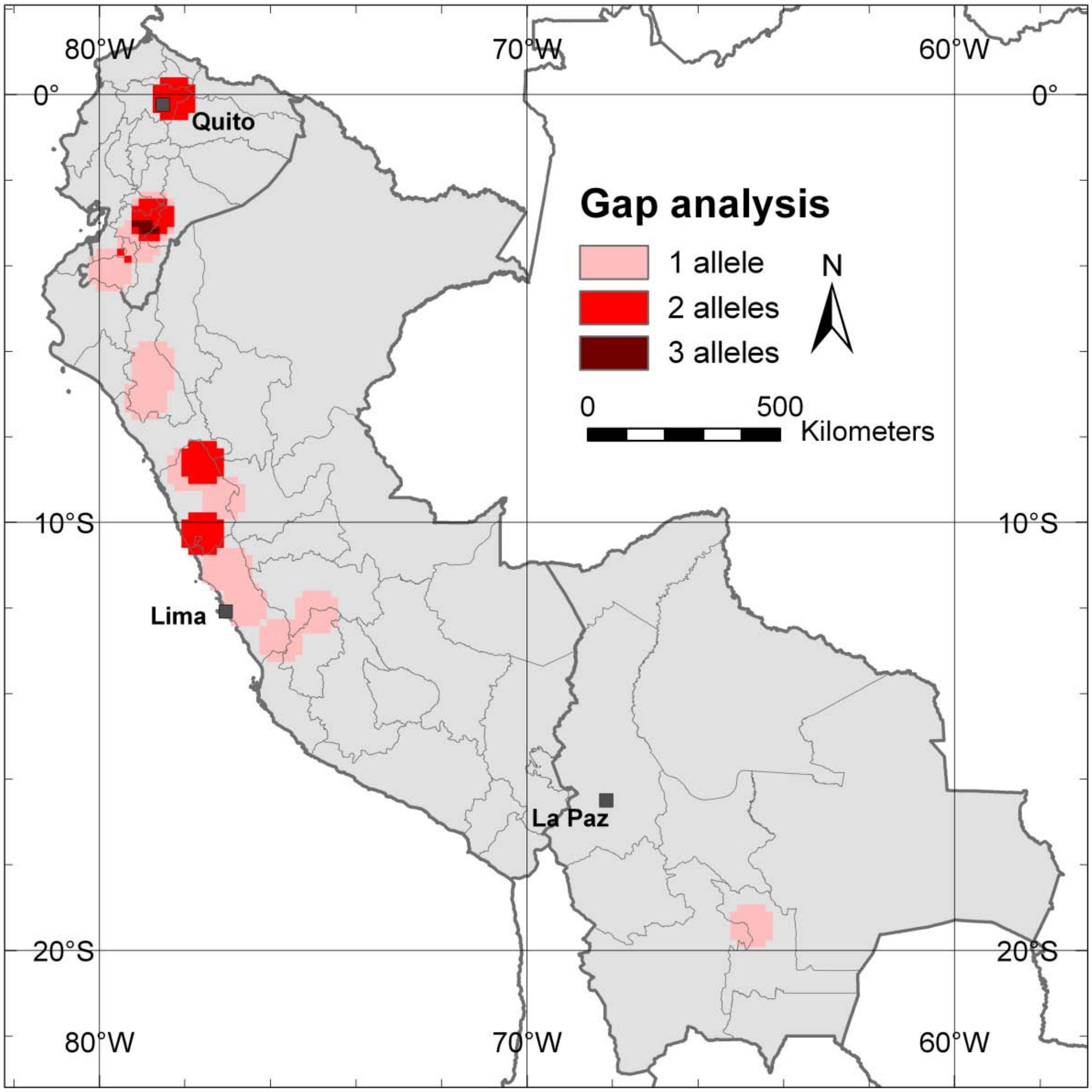


Figure 6.9 Richness analysis of alleles (eight alleles out of the total of 71 observed alleles) that are not found in any *ex situ* collection based on 10-minutes grid with a one-degree circular neighbourhood.

et al. 2005). AUC is a commonly used parameter in the validation of distribution models. Another measure of validation, the Kappa value, returned a value of 0.799 indicating the model performed even excellent (Fielding and Bell 1997).

In general, sampling covered most of the cherimoya-modelled distribution (Figure 6.10); 46 % of the modelled distribution area is covered by grid cells with 20 or more re-sampled trees (Figure 6.10, dark blue areas). In 24.5 % of the potential area of cherimoya occurrence less than 20 trees were re-sampled (light blue areas) whereas 29.5 % of the modelled range was not sampled (red areas) and can be considered sample gaps.

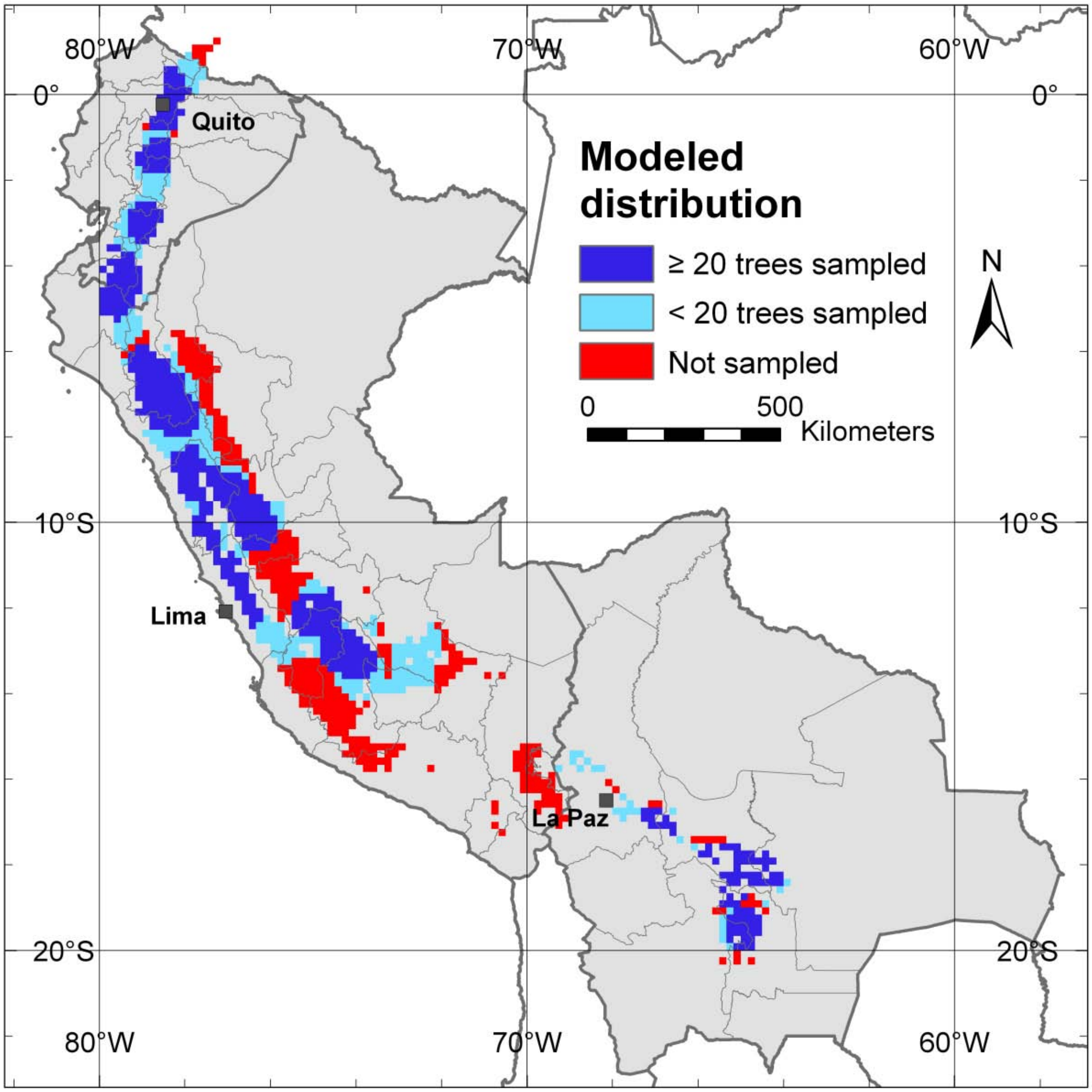


Figure 6.10 Areas of the modelled distribution in dark blue are covered by the 10-minutes grid cells with 20 or more trees after applying circular neighbourhood. Light blue areas of modelled distribution coincide with grid cells that contain less than 20 trees after re-sampling. Red areas indicate potential areas for cherimoya occurrence and cultivation that have not been sampled.

The largest sample gaps are located in northern Peru in the transition zone between the Peruvian Andes and the Amazon (in the Departments of San Martin and Amazonas) and in southern Peru (in the Departments of Junín, Pasco, Huancavelica, Ayacucho and Puno). The Andean-Amazon transition zone should reserve priority for future complementary cherimoya collection trips because it is adjacent to an area where already high levels of diversity have been found, i.e. Cajamarca Department in northern Peru.

Cherimoya was predicted by the distribution model to be absent from a significant area of southern Peru, indicating that the environmental conditions in substantial parts of that region are not suitable for cherimoya growth/or cultivation (Figure 6.10). This explains why no trees have been sampled in that area.

6.3 Discussion

Areas of high diversity in the cherimoya centre of origin

Our results are in line with a previous genetic study of the Spanish cherimoya collection that also distinguished populations in Ecuador and northern Peru from those in southern Peru (Escribano et al. 2007). They further corroborate results from isozyme markers that showed high genetic variation to be present in Peru and Ecuador (Perfectti and Pascual 2005). However, our study is based on a much higher number of samples. It, therefore, provides much more detail for prioritizing areas for *in situ* conservation and germplasm collection.

At allele level, our analysis confirms that, within our study area, the highest allelic richness together with the highest number of locally common alleles are found in southern Ecuador and northern Peru, i.e. the putative centre of origin of cherimoya. Northern Peru, and more specifically Cajamarca Department, shows the highest levels of genetic diversity.

The highest values of the fixation index, which is an indication of inbreeding, were found in Ecuador. Inbreeding may occur because of reduction and fragmentation of natural stands and cultivated areas, increasing the risk of allele loss, which eventually leads to genetic erosion (Lowe et al. 2005). Our results do not allow us to determine how much genetic erosion has taken place in Ecuador in comparison to Peru and Bolivia. However, high inbreeding values in Ecuador could explain why allelic richness is currently lower in this country than in northern Peru.

At population level, significant differences can be observed between the cherimoya germplasm present in the area with highest diversity (where genotypes belonging to cluster A are predominant), and genotypes found in areas with lower diversity, i.e. in southern Peru and Bolivia (represented by cluster B). Cluster A seems likely to represent material that is genetically closer to the ‘wild’ cherimoya type. No natural cherimoya stands have been observed in Bolivia, and this probably explains why no genotypes pertaining to cluster A have been recorded there. Cluster B probably corresponds to a gene pool that is genetically different from most of the wild or semi-domesticated cherimoya found in northern Peru and Ecuador and that could have formed the basis for cultivated Bolivian cherimoya. Looking at the areas with high cluster B dominance, Bolivian germplasm probably originates from southern Peru.

Although most early chroniclers and scientists proposed southern Ecuador and northern Peru to be cherimoya´s centre of origin (Popenoe 1921; National Research Council 1989b; Bonavia et al. 2004), the possibility of that area being a secondary centre of origin cannot be discarded. A diversity study similar to the one described in this study, but including cherimoya genotypes from Central America and Mexico, would shed light on the genetic variation across the complete pre-Columbian distribution range of cherimoya. It also would provide additional clues where to put the primary centre of origin and subsequent diversification of this species.

Having said that, our results clearly show that within cherimoya's Andean distribution range, northern Peru is a clear centre of diversity. This suggests that humans dispersed cherimoya from this area to Ecuador, the other parts of Peru and eventually to southern Bolivia where least genetic diversity was found. Cherimoya seeds have been identified from several pre-Columbian archaeological sites, including the site los Gavililanes from the so-called late Preceramic period (about 4,300 years ago) located in current Ancash, central-coastal Peru (Bonavia et al. 2004). Ceramic cherimoya-shaped vases were found in excavations of the Ecuadorian Valdivia culture that existed in the same period or earlier (5,500 – 3,600 years ago) (Wolters 1999). Considering that northern Peru is the centre of Andean cherimoya diversity, this implies that cherimoya cultivation started even earlier in this area.

Ex situ *and* in situ *conservation of cherimoya genetic resources in the Andean region*

Most alleles identified in our study are represented in one or more of the existing *ex situ* collections in Ecuador, Peru and Spain. Results obtained suggest that the highest priority for further collection should be Azuay Province in Ecuador, since cherimoya stands in this area harbour most alleles not yet included in genebanks. It is also one of the areas with the highest risk of allele loss because of the high observed levels of inbreeding, compared to other parts of the study area. An additional priority area for germplasm collection is the transition zone from the Andes to the Amazon in Peru (in the higher elevation areas of the Departments of San Martin and Amazonas), which was not sampled in this study. According to the distribution model there is a high probability of finding cherimoya stands in this region. The latter is probably also high in genetic diversity, because it is adjacent to the area with the highest diversity found in this study, i.e. Cajamarca Department in northern Peru.

A priority for conservation on-farm and in natural populations should be the Cajamarca Department, the area with the highest levels of genetic diversity. A second area of priority should be Loja Province in southern Ecuador, an area with a high number of locally common alleles. Both areas are assigned mostly to cluster A. Since trees assigned to cluster B have a particular allelic composition in comparison to trees predominantly grouped in cluster A, cluster B genotypes should also be considered in conservation activities. The part of Lima Department north of the Peruvian capital, which is assigned mostly to cluster B, could be prioritized for *in situ* conservation of genotypes from this cluster. In contrast to the low levels of allelic richness around Lima city in the southern part of Lima Department, the northern and more rural part of this Department contains a fair number of locally common alleles.

Long-term conservation of cherimoya genetic resources is far from guaranteed. As commercial prices for fruits can fluctuate, short-term incentives for farmers to maintain cherimoya as a profitable crop are uncertain. A decline in commercial interest may lead to replacement of cherimoya trees by other crops, increasing the risks of genetic erosion. An increase in commercial prices for cherimoya products will not necessarily promote conservation of existing genetic diversity. Indeed, in our study we found low levels of genetic diversity around the Peruvian capital, Lima. In this area the clonally propagated cultivar 'Cumbe' is widely cultivated because it currently fetches higher prices in the market.

A promising strategy to enhance on-farm conservation is through the promotion of seed or bud-for-grafting exchange between farmers (Tapia 2000). During the CHERLA project cherimoya fairs, which facilitate exchange of plant material, were organized in different areas of this study, including the Cajamarca and Piura Departments in Peru, Loja Province in Ecuador and various departments in Bolivia. This project aimed to promote sustainable cherimoya production systems in Latin America through the characterisation, conservation and use of local germplasm diversity (http://www.eelm.csic.es/proyecto/cherla/).

Seed and bud exchange can also be a way to conserve local races from unfavourable alterations in the local environment due to climate change, by re-distributing them in new areas with more suitable climate conditions (Mercer and Perales 2010). Another way to combine conservation of tree species genetic resources with their use could be through the establishment of local clonal seed orchards if and when adequate propagation techniques that enable the multiplication of clones, are made available as well (Cornelius et al. 2006; Ræbild et al. 2011). Cherimoya provides a good example, as demonstrated by the successful clonal propagation of the cultivar 'Cumbe' around the city of Lima.

Ideally, each area targeted for *in situ* conservation - where existing cherimoya stands and forest patches can evolve within the local environment - should be backed up by *ex situ* conservation of germplasm (which currently is the case for cherimoya genetic resources from the Andean region). Moreover they should be monitored periodically to assess dynamics in diversity use and risks of genetic erosion. *Ex situ* collections of fruit tree species often consist of living trees, such as illustrated by the cherimoya collections. This allows conservation of superior combinations of alleles that can be propagated vegetatively through grafting. Additional reasons for establishing field genebanks include the following: (1) many tropical and subtropical trees (including cherimoya) have seeds with recalcitrant or intermediate behaviour, so that they cannot be stored for long-term conservation; and (2) pollen, fruits and seeds can be collected continuously for characterization, evaluation and genetic improvement once trees have reached the reproductive stage. Nevertheless, the high costs for research institutions to maintain field genebanks of woody perennial species, can be a reason to downgrade *ex situ* collections and focus on *in situ* conservation (Van Leeuwen et al. 2005). In that case, it is important to screen the existing accessions through morphological, biochemical and/or molecular characterization to maximize in order to favour conservation of genetic diversity and potentially interesting functional attributes in a reduced collection (Frankel et al. 1995a). This approach has already been successfully used in the cherimoya collection la Mayora, Malaga, Spain (Escribano et al. 2008b).

Ex situ conservation may particularly be important for areas with materials that suffer from inbreeding -an indicator for high rates of allelic loss and genetic erosion- such as central Ecuador in the case of cherimoya. *In situ* conservation may be most successful in areas of high diversity where still low rates of inbreeding are observed such as in the cherimoya stands from northern Peru.

Use of GIS and molecular marker methods to enhance conservation and use of plant genetic resources

Despite the advances in new computational applications and the use of molecular tools, spatial analyses are still underutilized in efforts to plan for plant diversity conservation (Escudero et al. 2003). With respect to targeting collection sites and prioritizing the conservation of plant genetic resources, spatial analyses of diversity have been carried out mainly at the species level for crop gene pools (e.g. Hijmans and Spooner 2001; Jarvis et al. 2003; Scheldeman et al. 2007). Only a few studies have concentrated on mapping intra-specific diversity to enhance the conservation of genetic resources of specific crops and trees (e.g. Lowe et al. 2000; Kiambi et al. 2008). Kiambi et al. (2008) grouped samples using a grid to compare diversity between geographic areas of similar size, whereas Lowe et al. (2000) applied re-sampling to enable the calculation of diversity estimates with high degrees of confidence. However, these studies were carried out with fewer than 100 individuals per species. This limits the precision of the geospatial analysis that can be carried out over the geographic distribution range of species. Our analysis combines both techniques on a large dataset (1,504 trees), which can be conceptualized as a continuous distribution of plant individuals, in which each individual is connected to its representing neighbouring trees because they share the same seed system and/or breed with each other. Based on this concept, trees have been sampled in this study following a scattered distribution to calculate, across the Andean distribution range of cherimoya, several diversity estimates important to prioritize areas for conservation, including two recommended parameters: allelic richness (Petit et al. 1998) and the number of locally common alleles (Frankel et al. 1995a). Since the application of molecular tools is gradually becoming cheaper, intra-specific diversity studies with large datasets will probably become more common in the near future, allowing for similar studies on other tree species and annual crops.

Size of grid cells and width of the circular neighbourhood for this type of geospatial analysis depends on how many plant individuals have been collected across the landscape, and the minimum number of plant individuals that is considered sufficient to make confident estimates of genetic parameters per grid cell. Application of circular neighbourhood provides an effective way to decrease grid cell size. This facilitates detection of spatial patterns in genetic variation across an extensive distribution range. Re-sampling trees in the landscape, generates a high number of grid cells with a sufficient number of trees to make confident calculations of genetic parameters per grid cell. It also makes analyses less sensitive to changes in grid origin and enables the inclusion of isolated trees in the calculation of genetic parameters, i.e. together with their closest neighbouring trees.

Ideally, the sampling strategy for this type of analysis should be identified based on a pre-defined grid, aiming at measuring the same number of trees per grid cell. However, due to logistical constraints and because a species simply may be more abundant in some areas than in others, in practice, sampling will always remain sub-optimal to a certain degree. Of all genetic parameters that can be measured, allelic richness is most sensitive to uneven sampling. As a result, we have corrected sample size by rarefaction (Petit et al. 1998). Repeated subsampling of a minimum number of tree individuals per grid cell is another possibility to correct for sampling bias (Leberg 2002). This technique could also be used to correct other genetic parameters than allelic richness for sampling bias, such as expected heterozygosity, although these are less sensitive to uneven sampling (Lowe et al. 2004).

Our circular neighbourhood method is especially relevant when samples have been collected from single plants or farmers' fields across a landscape in contrast to the more 'classic' population genetic studies where in geographically separated populations a specific amount of plants is sampled (Lowe et al. 2004). The advantage of our method is that it allows to sample plant individuals across extensive environmental gradients and geographical ranges. Although ideally all samples should be collected according to a sampling strategy, another advantage of our methodology is that it allows to combine data from different collecting missions in one analysis. The mapping of cacao microsatellite diversity with data from different collecting missions such as presented by Thomas et al. (2012) is a good example of this. Many genebank-organized PGR collecting missions have sampled single plants or bulks from farmers' fields across environmental and geographical gradients. These then represent historic spatial diversity patterns of crops and other economic plant species. Our approach could thus be potentially interesting to establish a baseline genetic diversity map for these species because these collections are increasingly characterized with molecular data. Concepts such as fixation index and expected heterozygosity have been established under 'classic' population genetic concepts, which not necessarily have the same meaning in the circular neighbourhood approach. In our study, results from our fixation index map provided some interesting insights in geographic, genetic patterns such as clear-spread of the propagated local cultivar 'Cumbe' around Lima. Further research should be carried out on the use of these genetic parameters in the circular neighbourhood approach.

Given the sampling distribution in our study area and the fact that for the calculation of most genetic parameters, we maintained a minimum of 20 re-sampled trees per grid cell, we defined a cell size of 10 minutes and a circular neighbourhood with a diameter of one degree, which enabled us to detect spatial patterns of genetic variation at administrative level one (e.g. provinces or departments) in Ecuador, Peru and Bolivia. For studies of plant species, in which individuals are sampled more closely together, a more clumped distribution can be expected compared to our scattered sampling distribution and/or in lower densities across the landscape. In these cases larger grid cells and/or a larger width of circular neighbourhood could be applied, whilst always assuring a sufficient number of trees per grid cell. The overall resolution of the study will obviously be lower, but it still can provide useful information on geographical patterns of genetic variation across a species distribution range for PGR conservation and use.

Following Frankel et al. (1995b), we hypothesized that areas with high diversity as considered by neutral molecular markers, (like our microsatellite loci) have a high probability to contain genetic material that will also show diversity in functional traits, including traits of agronomic interest. Neutral markers that are generally not directly related to any specific function can correspond to population fitness, especially for out-breeding insect-pollinated species (Reed and Frankham 2003; Vranckx et al. 2011). Molecular markers are considered an appropriate indicator to quantify patterns and trends in the use and conservation of plant genetic resources because they can indicate the rates of kinship between varieties (Eaton et al. 2006). In the case of locally common alleles, they can evidence a long history of local human and natural selection (Frankel et al. 1995b; van de Wouw et al. 2010a) and be a proxy for effective population sizes (Petit et al. 1998). However, not necessarily neutral genetic variation is correlated to variation in quantitative, adaptive traits because they are shaped by different natural processes such as random drift and natural selection (Holderegger et al 2006). Neutral molecular marker surveys are practical for diversity studies. Direct measurement of traits in field trials on the other hand may be more desirable to evaluate genetic health and adaptive capacity of tree populations (Lowe et al. 2005). Nevertheless, molecular marker studies representative of the whole genome provide a cheaper and scientifically sounder alternative to assess the genetic resource status of tree species. In comparison to annual crops, perennial crop and tree field trials are particularly expensive because of the long generation times till first fruit set (Rajora and Mosseler 2001). Markers of DNA sequences related to phenotypic traits, including Expressed Sequence Tagged markers (EST) and markers in specific genes, could be of interest to include in geospatial analysis of patterns and trends in plant genetic resources. More and more markers are becoming available, especially for crops where sequencing programs have been performed or will be carried out in the near future. For a sister species of cherimoya, i.e. custard apple (*Annona squamosa* L.), for example, recently a gene has been described that plays an important role to form seedless fruits (Lora et al. 2011). However, these markers are less polymorphic than neutral ones, such as those that have been used in our study. So using neutral markers to study spatial patterns of genetic diversity is still necessary.

It is difficult to compare our results with those of Lowe et al. (2000) and Kiambi et al. (2008) because of the differences in methodology used. To compare molecular marker-based diversity studies on the same species, minimum standard sets of markers have been suggested (Van Damme et al. 2011). Standardization of methodologies in studies on different species would improve comparability of results. It also would facilitate meta-analyses, for example to better understand how well genetic diversity of tropical and subtropical tree species is conserved on-farm and in protected areas.

In our study, we only examined spatial patterns of genetic variation without relating them to other spatial attributes. GIS can also be used to link genetic data to available spatial information relevant to conservation of plant genetic resources. GIS can thus be uses to reveal both short-term threats (such as accessibility) and long-term threats (such as climate change). With this type of analysis, hotspots of diversity under threat could be identified following Myers et al. (2000). However, instead of looking at species level, this could be done at intra-specific level, to ensure conservation of priority populations of specific crops and useful tree species. Spatial information on patterns and characteristics of human societies can be used to understand the drivers behind threats. In a study on changes in cassava diversity in the Peruvian Amazon, GIS was used to correlate cassava diversity data with biotic and socio-economic spatial data to identify possible drivers behind diversity and genetic erosion (Willemen et al. 2007). This would be useful information in the development of adequate policies and measures to promote *in situ* conservation of plant genetic resources on farms and in natural populations.

6.4 Methods

Sampling and SSR analysis

A total of 1,504 cherimoya accessions have been analysed in this study, 395 from Bolivia, 351 from Ecuador and 758 from Peru. DNA was extracted from young leaves following Viruel et al. (2004). Based on polymorphism, a set of nine SSRs was selected from those previously developed in cherimoya (Escribano et al. 2008a). A 15 µl of reaction solution containing 16 mm (NH4)2SO4, 67 mm Tris-ClH pH 8.8, 0.01 % Tween20, 2 mm MgCl2, 0.1 mm each dNTP, 0.4 µm each primer, 25 ng genomic DNA and 0.5 units of BioTaq^{TM} DNA polymerase (Bioline, London, UK) was used for amplification on an I-cycler (Bio-Rad Laboratories, Hercules, CA, USA) thermocycler using the following temperature profile: an initial step of 1 min at 94 ºC, 35 cycles of 30 s at 94 ºC, 30 s at 45 ºC-55 ºC and 1 min at 72 ºC, and a final step of 5 min at 72 ºC. Forward primers were labelled with a fluorescent dye on the 5' end. PCR products were analysed by capillary electrophoresis in a CEQ^{TM} 8000 capillary DNA analysis system (Beckman Coulter, Fullerton, California, USA). Samples were denaturalized at 90 ºC during 120 s, injected at 2.0 kv for 30 s and separated at 6.0 kv during 35 min. Each reaction was repeated twice and the Spanish cultivar 'Fino de Jete' was used as control in each run to ensure size accuracy and to minimize run-to-run variation.

Data cleaning

The coordinates of the respective tree locations were checked in DIVA-GIS (www.diva-gis.org) on erroneous points based on passport data at administrative level one (e.g. departments, provinces) with a buffer of 20 minutes (approximately 30 km). They were also checked on outliers based on climate data derived from the Worldclim data set (Hijmans et al. 2005b) two or more of the 19 bioclim variables according to the reverse jack-knife method (Chapman 2005). Based on these analyses, two points were excluded. The cleaned dataset thus included microsatellite data of 1,504 georeferenced trees. Taking into account that nine SSR markers were analysed, this results in a total of 27,072 georeferenced alleles.

Geospatial analysis – Circular neighbourhood

Grids for all genetic parameters were generated in DIVA-GIS and are based on a grid with a cell size of 10 minutes (which corresponds to approximate 18 km in the study area). On these rasters we applied a circular neighbourhood with a diameter of one degree (corresponding to approximate 111 km) constructed in Excel. The circular neighbourhood is used to re-sample the allelic composition of a single tree to all surrounding grid cells. In this case, this meant 32 cells with a size of 10 minutes, within a diameter of one degree around its location. In this way, the allelic composition of each sampled tree is representative for the area within the defined buffer zone. Applying the circular neighbourhood re-sampling technique resulted in a total dataset of 48,128 trees and 866,304 alleles.

Geospatial analysis – alpha diversity

After applying circular neighbourhood to all trees, genetic parameters were calculated in GenAlEx per 10-minutes grid cell, for all trees present in each cell after re-sampling. Genetic parameters included average number of alleles per locus (*Na*), number of locally common alleles per locus (alleles occurring with a frequency higher than 5 % in 25 % or less of grid cells), average expected heterozygosity per locus (*He*), fixation index (*F*) and genetic distance (*GD*) (see Peakall and Smouse 2006). *Na* and the number of locally common alleles per locus were presented for all grid cells with trees included. *Na* was corrected by rarefaction to a minimum sample size of 20 trees per cell with HP-RARE software (see Kalinowski 2005). Consequently, this parameter was only calculated for grid cells with 20 or more re-sampled trees. This minimum sample size was also used as a threshold of the number of trees per grid cell to get interpretable results for the parameters *He*, *F* and *GD*. The latter parameter was used to calculate distance in allelic composition of each cherimoya genotype to the commercial variety 'Cumbe'. This value was calculated in GenAlEx using the *GD* option for co-dominant markers (see Smouse and Peakall 1999). Final *GD* value per grid cell was the average *GD* for all re-sampled trees present in each cell. The reference tree was the accession 'Cumbe' from the Spanish cherimoya genebank in Malaga.

Geospatial analysis - beta diversity

Population structure was defined by running the Structure software (see Pritchard et al. 2000) on all 1,504 samples applying a 10,000 burn-in period, 10,000 Markov Chain Monte Carlo (MCMC) repetitions after burn-in, and 20 iterations. Optimal K was selected after Evanno et al. (2005) by running Structure for K values between one and 10, and defining the final number of clusters where ΔK value was highest. This was at K=2, hence a map was developed for these two clusters, which we named respectively A and B. We used the probabilities of each tree belonging to cluster A and B to visualize the clusters on a map. Mapping of probabilities was done based on the average value of all trees per 10-minutes cell for those grid cells with 20 or more re-sampled trees after applying the one-degree circular neighbourhood.

Geospatial analysis - ex situ *conservation status*

The private alleles function in GenAlEx (PAS) was used to identify the alleles exclusively found in trees that were sampled *in situ*. To visualize patterns in these alleles that are not included in any genebank, a point-to-grid richness analysis, using a 10-minutes grid, was carried out in DIVA-GIS based on the one-degree circular neighbourhood re-sampled tree grid.

Geospatial analysis - distribution modelling

To identify how well the sampling covered the Andean distribution range of cherimoya, and thus to identify potential collection gaps, we modelled the distribution (presence only) of cherimoya in the study area using the distribution modelling program Maxent (see Phillips et al 2006; Elith et al. 2011). With this technique, potential distribution areas are identified as areas where similar environmental conditions prevail as those at sites where the species has already been observed. The data required to identify these areas include species presence points as well as layers of environmental variables covering the study area. Maxent is an EEM tool for which the applied algorithm has been evaluated as performing very well, in comparison to other EEM software (Elith et al. 2006; Hernandez et al. 2006). Therefore, it was selected for this study's distribution modelling analysis. The coordinates in the passport data of the sampled trees were used for the presence point input. For environmental layer input, we used the 10-minutes grids of 19 bioclimatic variables (see Busby 1991), derived from the Worldclim dataset (Hijmans et al. 2005b).

The modelled distribution area was restricted using the 10 percentile training presence threshold, which indicates the probability value at which 10 % of the presence points falls outsides the potential area. The modelled distribution was generated in Maxent with 80 % of points (training data). It was subsequently cross-validated in DIVA-GIS with 20 % of the remaining tree observations (test data). Besides 20 % of presence points, test data included randomly generated points in 0.1 times the bounding box of the presence points as a proxy for absence points (5 times the number of presence points). Based on the cross-validation, Area Under Curve (AUC) and Kappa value were calculated in DIVA-GIS as measures of model performance.

All maps were edited in ArcMap.

Discussion

7. Usefulness of spatial diversity and distribution analysis for plant genetic resources conservation and germplasm collecting

In this thesis, I aimed to test the following hypothesis: spatial analysis of plant diversity and distribution can clearly detect geographic inter-specific and intra-specific diversity patterns, which allows to prioritize populations and geographic areas that should be considered for *in situ* conservation and germplasm collecting. In the first two chapters of this thesis, general concepts of geographic plant diversity and distribution analysis and methods were explained. In addition, practical recommendations were provided to use Geographic Information Systems (GIS) tools for plant genetic resources (PGR) *in situ* conservation and germplasm collecting.

GIS tools always return a result independent of the data quality of which an attractive map can be developed. Therefore another point highlighted in chapter 1 on ecogeographic diversity and distribution analysis is the importance of data preparation and quality. These are the basis of all sound analyses. Special emphasis was given on the application of molecular marker methods for PGR characterization of *in situ* samples to identify hotspots of intra-specific diversity. This is easier and less costly than morphological characterization of intra-specific diversity in *ex situ* common-garden experiments.

The geographic distribution patterns for many plant species in the tropics and subtropics are poorly known, that's why often Ecological Envelope Modelling (EEM) is used to make inference in species' potential distribution ranges. Although botanists and other professionals in the field are often cautious about the results and applicability of EEM, the results from the expert evaluation exercise in chapter 3 indicate that many professionals are fairly positive about the use of EEM for *in situ* conservation planning. These results support the hypothesis of this thesis, and suggest that such analyses can support local professionals in their planning work of managing and conserving plant genetic resources. Interestingly, the most knowledgeable specialists, i.e. the ones that followed best the consensus of the expert groups, tended to be more positive in their model appreciations than specialists in their group that agreed less to the consensus. This affirms the fairly positive feedback that species specialists provided in general. However, the precision of EEM may remain low to support several field activities of the experts. Further development of distribution modelling techniques to provide support for more local conditions would therefore be helpful. A key point is to increase the availability of more detailed geospatial environmental layers.

The results of the case studies in chapter 4 and 6 on respectively geographic distribution patterns of endemic wild potato (*Solanum* spp.) relatives richness in Bolivia and cherimoya allelic richness in the Andean region provides further support to the hypothesis of this thesis. These case studies show that spatial analyses of plant diversity and distribution can clearly detect geographic patterns of inter- and intra-specific diversity to prioritize areas for conservation. In the case of endemic wild potato species occurrence in Bolivia, a clear hotspot of endemic diversity could be identified in the inter-Andean valleys of south-central Bolivia. Northern Peru is a hotspot of genetic cherimoya (*Annona cherimola* Mill.) diversity in the Andean distribution range of this Neotropical fruit tree species. The study on cherimoya demonstrates that clear spatial patterns of intra-specific diversity can be detected relatively easy when standardized methods are used across an extensive distribution range with sufficient samples. In both cases the results can be used to provide recommendations to prioritize conservation actions at the level of departments or provinces and also within these administrative units.

Although several existing peach palm (*Bactris gasipaes* Kunth) diversity studies have been carried out, no clear results were found yet about species origin and centres of domestication. The latter studies deal with a limited number of individuals, populations and/or use different molecular markers and methods in existing studies. This makes comparison difficult.

Of course, the diversity, distribution and threat analyses presented in this thesis can be applied to other species as well. In this sense, the potato wild relatives, cherimoya and peach palm are model taxa. The analyses are now being used to assess the diversity and conservation status of 100 socio-economically important tree species native to Latin America. This is done in MAPFORGEN, an online platform that has been established in collaboration with species specialists (www.mapforgen.org). More specifically, the circular neighbourhood analysis that was used to better understand spatial patterns of intra-specific diversity presented in the cherimoya study, has now been applied in studies of several other plant species. The list include cacao (*Theobroma cacao* L.) (Thomas et al. 2012), *Cedrela balansae* C.DC. (Soldati et al. 2013), wild barley (*Hordeum vulgare* var. *spontaneum*) (Russell et al. *in prep.*) and *Nothofagus* spp. (Azpilicueta et al. 2013).

Many professionals working with plant genetic resources and in/or in the field of economic botany may opt to apply themselves geospatial analyses but do not necessarily have the required experience. Promoting capacity building and training materials could bridge this gap. Parallel to the research for this thesis, a manual on plant diversity and distribution was developed. This manual responds to the increasing demand of professionals working with plant genetic resources such as botanists, agronomists and ecologists for this type of analyses (Scheldeman and van Zonneveld 2010). The tutorial presents exercises to practise geospatial analyses derived from existing publications (Scheldeman et al. 2007; van Zonneveld et al. 2009b). The latter formed the basis for the more elaborated concepts and studies presented in this thesis. The training manual is intended for self-learning. To date, these exercises have been used in more than 20 courses, mostly in Latin America, but also in Europe and Africa. The manual is recommended by the Global Biodiversity Information Facility (GBIF) for diversity and distribution analysis (http://www.gbif.org/orc/?doc_id=2917&l=en). It can be freely accessed online in English, Spanish and French to facilitate its use for professionals in different parts of the world.

8. Drivers of diversity patterns

In this dissertation, I focused on *where* areas of high diversity are located and what type of diversity indicators and geospatial techniques are appropriate to detect them. Among different tools, species distribution models have been very popular tools in ecological and biogeographic studies. They are used to estimate species presence and levels of species diversity in areas that have not been surveyed yet. Such models also help to better understand how biophysical, ecological and evolutionary factors shape species distributions and diversity at different time and spatial scales. This knowledge helps to predict the impact of possible changes in environmental factors on existing plant distribution and diversity.

In chapter 3, 4 and 6 of this dissertation a commonly used type of distribution modelling was applied (presence-only Environmental Envelop Modelling [EEM] with the use of Maxent) to predict species presence and species diversity of a specific gene pool. The distribution of natural populations was modelled for the timber tree species *Nothofagus alpina* (Poepp. & Endl.) Oerst. and *Cedrela odorata* L. that were used as model species in chapter 3 and the wild potato relatives in chapter 4. The distribution of both natural populations and cultivated populations was predicted for peach palm and cherimoya within their pre-Columbian distribution range. Most Brazil nut (*Bertholletia excelsa* Humb. & Bonpl.) stands are considered not to be cultivated. However, pre-Columbian human cultures are thought to have actively promoted regeneration and dispersal of this species (Sheppard Jr and Ward 2011).

In general, EEM is used to understand and predict the distribution of wild species, and not of cultivated species, as was done with peach palm, Brazil nut and cherimoya in this dissertation. This is because EEM has been traditionally developed in ecology to understand the relationships between species and their environment. Nevertheless, the technique has also been used to model the distribution of cultivated species that are locally and regionally important. Since no sufficient productivity data exist for species such as Brazil nut, peach palm (*Bactris gasipaes* Kunth) and cherimoya (*Annona cherimola* Mill.) to develop crop models, EEM can be used as an alternative to estimate suitable growing areas (Bowe and Haq 2010). Other examples from the Americas of modelled distributions of cultivated plant species include jocote (*Spondias purpurea* L.) (Miller and Knouft 2006), highland papayas (*Vasconcellea* spp.) (Scheldeman et al. 2007) and agaves (*Polianthes* spp.) (Solano and Feria 2007). Moreover, many of these species in the Americas are traditionally cultivated and/or maintained in semi-natural habitats (e.g. Clement 1999; Scheldeman et al. 2003). This suggests that they are adapted to specific environmental conditions and are not intensively managed. Nevertheless, the niche of cultivated plant species can expected to be wider compared to the climate ranges in which wild species populations occur because of domestication for adaptation to different types of growing conditions, management practises and reduction of competence (Miller and Knouft 2006).

As distribution of cultivated plants is determined by both cultural as environmental factors, an interesting study topic would be to investigate how much the modelling of cultivated plant species distributions will improve when cultural variables are included. Eco-cultural niche modelling approaches (ECNM) have already been applied to better understand the distribution of Pleistocene cultures and the environments in which they lived (Banks et al. 2011). The explanatory variables used in these studies include downscaled layers of past climate projections and the localities of archaeological findings (Banks et al. 2011). Envelope modelling could test the predictive power of presence maps of archaeological crop residues as an explanatory variable of cultivated plant species distribution.

Localities of archaeological plant remains have already been used to validate crop dispersal routes (van Etten and Hijmans 2010). However in some culturally rich areas, such as the Amazon, historic cultivated plant residues are difficult to find (Pearsall 1992). This leads to bias in modelling results. Complementary or alternative cultural variables for cultivated plant species niche modelling could be distance to historical human routes (Levis et al. 2012) and linguistic diversity (Gorenflo et al. 2012). How well such cultural factors improve distribution modelling of cultivated plants remains to be tested in further studies.

Geographic patterns of plant diversity and distribution are shaped by different drivers according to the spatial scale of analysis. In this thesis, layers of climatic variables were used as input for EEM at regional and national scale. At these large scales, climatic variables are important factors to explain geographic patterns of natural species diversity and distribution (Pearson and Dawson 2003; Field et al. 2008). Other responsible factors for natural plant diversity patterns and distribution ranges at regional scales are radiation and latitude-related variables (Willis and Whittaker 2002) such as photoperiodicity (Vavilov 1992c) and species colonization towards the poles after the last glacial period (Willis and Whittaker 2002; Svenning and Skov 2007). Environmental factors like soils and elevation play a bigger role in shaping plant distribution and diversity at smaller spatial scales and shorter time scales whereas biotic and abiotic interactions drive species diversity and distribution in local vegetation communities (Willis and Whittaker 2002; Pearson and Dawson 2003).

At intra-specific level, geographic patterns of molecular diversity can help to identify species dispersal routes at regional level. In the case of wild species, different types of molecular markers can detect processes of reestablishment and colonization from refugia after the glacial populations (Newton et al. 2001; Petit et al. 2003). Hotspots of genetic diversity can be explained by environmental and ecological characteristics of glacial refugia and associated demographic processes that assured sufficiently high effective population sizes in those past periods (Taberlet et al. 2012). And also human-mediated dispersal routes can be detected. The results of the cherimoya molecular diversity study in chapter 6, for example, provide insights into cherimoya dispersal in the Andes.

Both the geographic patterns of taxonomic diversity of the wild potato (*Solanum* spp.) relatives' gene pool (chapter 4) and molecular cherimoya diversity (chapter 6) were mapped at regional level and are clearly clumped in hotspots of diversity within their respective distribution ranges. However, the drivers that explain the diversity patterns in these two examples are different from each other.

The occurrence and evolution of potato wild relatives have been shaped by natural processes during hundreds, thousands and millions of years. Cherimoya is a cultivated species in its incipient phase of domestication (National Research Council 1989b). The distribution of cherimoya genetic diversity follows a dispersal pattern from a centre of crop diversity according to the concepts of crop geography and centres of origin of cultivated plants proposed by Vavilov (1887-1943). His concepts have been improved continuously since then (Zeven and de Wet 1982; Frankel et al. 1995a; Vavilov 1992b; van Etten and Hijmans 2010). According to this theory, genetic diversity is expected to decline from a crop's centre of diversity towards the boundaries of the distribution ranges. This is because during human dispersal, crops and varieties go through a genetic bottleneck each time when they are introduced in new areas (van Etten and Hijmans 2010). Human-mediated crop dispersal is hypothesized to have started about 13,000 years ago when the plant cultivation and domestication began in different parts of the world, including in Mesoamerica and the Andean region (Mannion 1999; Purugganan and Fuller 2009; Meyer et al. 2012). The oldest evidences of cherimoya cultivation in south America are found in the coastal areas of Ecuador and Central Peru and are dated 5,500 – 3,600 years ago (Wolters 1999; Bonavia et al. 2004).

In the case of cherimoya in its Andean distribution range, humans probably spread the fruit species from the hotspot of diversity in northern Peru towards other Andean parts where current cherimoya stands harbour lower *alpha* diversity. These areas include Ecuador, southern Peru and Bolivia. Secondly, the Andean distribution of cherimoya seems to follow the isolation by-distance model of Wright (1943). According to this model geographically more distant plants are also genetically more distinct because their ancestors have had little opportunity to interbreed with each other. This results in lower kinship rates between geographically more isolated plant individuals (van Etten and Hijmans 2010).

The fact that evidences of cherimoya cultivation are found 5,500 – 3,600 years ago in Ecuador and Central Peru, suggests that in northern Peru cherimoya cultivation and possible domestication started even earlier. Whether northern Peru is the primary or a secondary centre of cherimoya diversity still needs to be confirmed by comparing cherimoya diversity across its complete pre-Columbian distribution range including the Andes, Central America and Mexico. Such a study also would provide more insight in north-south human-driven exchange in pre-Columbian times between North, Central and South America.

Some authors have argued that north-south crop dispersal has been limited due to geographic constraints especially when compared to east-west crop dispersal such as in Eurasia and crop dispersion from the centres of crop domestication in China and the Fertile Crescent (Diamond 2002). North-south dispersal may also have been restricted because adaptation of crops and their varieties in new areas at respectively higher or lower latitudes can be limited due to differences in day length patterns. Crops and plant species may face difficulties to adapt to new light conditions (Vavilov 1992c). Changes in these conditions influence photosynthetic activity and plant phenology (Vavilov 1992c). However, the results from the molecular cherimoya study taken together with results from existing molecular potato and maize (*Zea mays* L.) diversity studies, demonstrate that American crops have been actively dispersed by humans in pre-Columbian times across the Andes from north to south, and also between Mexico, Central America and South America (Spooner et al. 2005; van Etten and Hijmans 2010).

As already mentioned before, cherimoya molecular diversity declines from its northern Peruvian hotspot towards northern Ecuador and southern Peru and Bolivia. This suggests active human dispersion in latitudinal directions across the Andes. Archaeological findings of ancient cherimoya seeds and ceramics in coastal Peru and Ecuador suggest that this happened in pre-Columbian times (Wolters 1999; Bonavia et al. 2004). Potato molecular studies with AFLP markers indicate a single domestication of cultivated potato from its wild relatives in southern Peru, in contrast to a previous hypothesis of multiple domestication events. From there, the species is believed to have spread in pre-Columbian times across the whole Andes; towards southern Chile and northwards up to western Venezuela (Spooner et al. 2005). Maize, finally, has spread from north to south from Mesoamerica towards different ecological zones in South America in pre-Columbian times (Vigouroux et al. 2008; van Etten and Hijmans 2010).

How fast crop and variety adaptation can occur across north-south dispersal axes requires further study, for example by evaluation of landraces in multi-location trials at different latitudes. This will help to better understand pre-Columbian crop dispersal and to identify more precisely corresponding centres of crop diversity. These centres of diversity are an invaluable source for plant breeding to be prioritized for *in situ* conservation of plant genetic resources and monitoring of the use of this diversity. Field experiments also provide knowledge on the potential of crops and varieties to shift north or southwards as a strategy to adapt agricultural systems to climate change. This climate change adaptation strategy consists of the introduction of crops and varieties from areas with already high temperatures and specific precipitation patterns that resemble the new local climate conditions. However, these ‘new’ plant species and varieties may require still some time to adapt in their new sites to local solar radiation regimes and length of growing season (and probably also to many other novel environmental conditions). Local farm selection and participative breeding programs can help to further adapt introduced crops and varieties to these unique local conditions.

It is well possible that mountain ranges, despite their topographical constraints for human movement and transport, facilitated human-mediated crop dispersion. At all latitudes of a mountain range, a wide range of environments can be found within short geographic distances. Another aspect of the same principle is that high environmental variability facilitates the establishment of diverse, resilient and sustainable agricultural production systems (Diamond 2002). The Inka for example, cultivated a wide variety of crops at different altitudes and Andean slopes within relatively short geographic distances (Mann 2005). Crop diversification across the Andean landscape enabled them to spread risk in their agricultural system (Mann 2005).

So, in each new location within a mountainous range, relatively easy, suitable environments can be found that resemble the growing factors from a crop's or variety's origin. Optimal temperature conditions can be found within short geographic distances by moving up or down mountain hills, and radiation patterns can be regulated by changing the position towards the sun, to name a few examples. Such straightforward adaptations are not possible when crops and varieties are moved across north-south axes of plain areas. In these areas, it is much easier to disperse crops and varieties in east-west direction (see Diamond 2002). Landscape models of crop dispersal should be developed to test whether mountainous areas indeed have facilitated human-mediated crop dispersal and to which extent.

In addition to human-mediated dispersal across the Andean mountain ranges, extensive coastal shipping happened across the pacific littoral of South America and Mesoamerica and between these regions during pre-Columbian times (Wolters 2001). For example, there are strong evidences that about 3,200 years ago, the Ecuadorean chorrera culture and Mesoamerican Olmecs, had established trade contacts (Wolters 2001).

Sea transport between these two regions has had led to early exchange of several crops including staples (maize, cassava [*Manihot esculenta* Crantz], common bean (*Phaseolus vulgaris* L.), peanut [*Arachis hypogaea* L.], sweet potato [*Ipomoea batatas* (L.) Lam.]); fruit species (cherimoya, avocado [*Persea americana* L.], cacao [*Theobroma cacao* L.]); horticultural crops (pumpkin [*Cucurbita pepo*], one of the five domesticated chili peppers [*Capsicum annuum* L.], tobacco [*Nicotiana* spp.]) and Mexican cotton (*Gossypium hirsutum* L.) (Wolters 2001). Certainly, this exchange had played an important role in further diversification of local production systems in both areas. It also contributed to broadening the gene pools of the transferred crops. For several crops the direction of crop exchange has been clear such as for cacao, peanuts and cassava. These crops were brought from South America to Mesoamerica. In the opposite direction, *C. annuum* and Mexican cotton were brought to South America. For many others crops such as sweet potato and cherimoya this has yet to be clarified. As mentioned before in this chapter, molecular analysis of the complete pre-Columbian distribution range of cherimoya, as well as for other crops can help to trace back historic dispersal routes and directions through detection of diversity hotspots, bottlenecks and kinship between materials from different areas. There are several cultural, biophysical, ecological and evolutionary factors that would have been responsible for the geographic patterns of cultivated plant diversity and dispersal. I will go further into detail about possible factors that drive cultivated plant diversity in the next chapter on future challenges.

Fascinatingly, the endemic, wild potato hotspot in Bolivia coincides with one of the four wild chili pepper relatives (*Capsicum* spp.) hotspots in the Americas to which the *Capsicum* genus is native. *Capsicum* belongs to the Solanaceae. This is the same botanical family to which the wild potato species belong (Spooner and Salas 2006).

In fact, the putative centre of origin of *Capsicum* is hypothesized to be in the inter-Andean valleys in south-central Bolivia because of the current high *Capsicum* diversity present (McLeod et al. 1982; Eshbaugh 2012). The latter includes the phylogenetically most-ancient existing wild chili pepper, *C. chacoense* Hunz. (McLeod et al. 1982; Eshbaugh 2012). The other *Capsicum* hotspots are the dry coastal forests in northern Peru and southern Ecuador, the Atlantic coastal forest in Brazil, and the dry forest in the Caatinga region of Brazil [Moscone et al. (2007) after Hunziker et al. (2001)].

Looking at a continental scale, south-central Bolivia is one of six wild potato relative hotspots in the Americas. One wild potato hotspot is located in the central Mexican highlands whereas the others can be found across the Andes (Hijmans et al. 2002). The latter authors defined central Bolivia as one of the six centres of wild potato relative diversity. Our results in chapter 4 indicate that a centre of endemic wild potato diversity occurs more specifically in south-central Bolivia, which is thus also the principal wild chili pepper hotspot in this country. Urgent conservation measures are required to maintain the hotspot of these two crop gene pools because this area is highly threatened by human disturbance including livestock activities pressure and increased human accessibility.

The hotspots of *Capsicum* fall into Neotropically Seasonally Dry Tropical Forests (STDF). SDTF are scattered across Latin America and can be found between 500 and 2,500 m (Pennington et al. 2000; 2010). Many of these forests are restricted in their distribution and are highly threatened by agricultural expansion (Pennington et al. 2010). They include inter-Andean valleys, such in south-central Bolivia, and are isolated from each other by more humid vegetation types (Pennington et al. 2000). The disjunctive vegetation distribution of SDTF can explain well why at continental scale a few wild chili pepper hotspots with high concentration of endemic species occur, and then a number of extensive distribution areas with low species diversity. Currently restricted STDF are hypothesized to have been more wide-spread during several colder and drier periods in Pleistocene and pre-Pleistocene epochs (Pennington et al. 2004). However, how STDF distributions looked like in the past and when and how they became isolated is still open to debate (Mayle 2004; Pennington et al. 2004).

Most wild potato species endemic to Bolivia occur in warm temperature climates with dry winters and warm summers, which correspond to mid-elevation forests and SDTF vegetation. However, in contrast to wild chili peppers, most wild potato species can be found in high-Andean vegetation above the upper elevation limit of SDTF of about 2,500 m. Therefore, it can be anticipated that the drivers for the current distribution of wild potato diversity are different from the ones that are responsible for current geographic patterns of wild *Capsicum* diversity.

Nevertheless, the different hotspots of both wild chili pepper and include many endemic species that not occur in other areas of high chili pepper and potato diversity [Moscone et al. (2007) after Hunziker et al. 2001; Hijmans et al. 2002]. This suggests that these nuclei of diversity have been isolated a very long a time ago and persisted over that time. Phylogenetic studies with other plant species indicate that the flora in some close-by inter-Andean have already have become isolated from each other at least five millions of years ago (Pennington et al. 2010).

Interestingly, a disjunct distribution has been observed for one wild potato species (Simon et al. 2011). This species has been observed in Central America and in Bolivia but surprisingly not in other parts of South America. This distribution pattern is similar to the scattered distribution ranges of plant species in SDTF (Pennington et al. 2000). Again, this suggests that the ecosystems that are inhabited by wild potato species may have been fragmented and isolated a long time ago. Processes that drive geographic patterns and speciation in gene pools at continental scale take millions of years and included past aridification events, quaternary glacial and interglacial cycles, and mountain-building episodes such as the Tertiary uplift of the Andes (Willis and Whittaker 2002).

Further research should be carried out to identify whether disjunctly distributed vegetation types such as SDTF and mid-elevation Andean forests harbour exceptionally high CWR diversity compared to others. If this is true, under-sampled areas of these vegetation types can be prioritized for further germplasm collecting of CWR. Many unknown CWR that contain interesting traits for plant crop breeding may still occur in these areas. In the case of chili peppers for example, regularly, new species are being described (Eshbaugh 2012). It is also expected that in wild potato relative hotspots many unknown species are yet to be found (Hijmans and Spooner 2001).

9. Further challenges

Combination of GIS and statistical software

Although GIS packages are being constantly improved, and spatial diversity and distribution analysis software (including DIVA-GIS) now include a good range of statistical tools, further integration of more advanced and specific statistical analyses and packages should be envisaged. Fortunately, new packages of R statistics have been developed for genetic and geospatial analysis (e.g. Jombart 2012; Hijmans et al. 2012; Hijmans and van Etten 2012). Additional efforts could be made to incorporate the analyses carried out in genetic statistical programmes (e.g. Adegenet, GenAlEx and STRUCTURE) into a GIS environment, allowing a more immediate and powerful graphical display of the results of studies of intra-specific genetic diversity.

Data sharing and standardization

An overarching technical challenge is enabling open access to existing and emerging sources of environmental and biological, as well as socio-economic, data by developing clear data-sharing rules; common formats for interoperability across software and hardware; open-source tools for data conversion, visualization and analysis; and automated dataset preparation. Improving access and integration of data will greatly facilitate the interdisciplinary approach required in biodiversity research, while supporting related policy-making initiatives (Canhos et al. 2004).

As already mentioned, the use of standardized sets of molecular markers, is becoming increasingly necessary in order to allow comparisons among the growing body of data on molecular diversity being generated worldwide. This is particularly true for major crops and increasingly for other species as well. Standardized characterization sets, which basically perform as descriptor lists at the morphological level, already exist. An example of standardized molecular characterization protocols are the Generation Challenge Programme (GCP) microsatellite (SSR) kits for 11 crops (among which are wheat [*Triticum* spp.], rice [*Oryza sativa* L.], maize [*Zea mays* L.], potato [*Solanum tuberosum* L.], sorghum [*Sorghum bicolor* L.], chickpea [*Cicer arietinum* L.], common bean [*Phaseolus vulgaris* L.]) (http://s2.generationcp.org/gcp-tmm/web). Particularly if and when the application of molecular marker methods becomes standardized, open access databases for molecular characterization data could be created and made easily accessible to users, complementing the information on species distribution data found in databases such as the Global Biodiversity Information Facility (GBIF) and Genesys (see chapter 2). This combined information would help the identification of hotspots of intra-specific diversity, informing *in situ* conservation strategies, directing collection missions and sampling aimed at material carrying specific traits, as well as detailed monitoring of economic plant diversity. Standardization would be most feasible for markers that have already been developed and are widely applied to non-model species to enable wide application of standards; these are mostly markers that measure neutral diversity.

Evaluation of phenotypic diversity

Several authors stress the relevance of characterization of phenotypic variation (Holderegger et al. 2006; Cleveland and Soleri 2007). This data is important to estimate the evolutionary potential of populations. However, it would be very expensive to phenotypically characterize and evaluate all populations within a species range.

To measure intra-specific diversity, I focused in this dissertation on molecular diversity because of the practical advantages in sampling and characterization. In the case study on cherimoya (chapter 6), neutral microsatellite markers were used to (1) identify centres of high *alpha* diversity where there is a high likelihood to find many functional traits including unknown ones; and (2) detect different genetic populations (*beta* diversity) which have followed different paths of selection and evolution. In this way, a subset of diversity can be identified that is representative for the whole diversity. This subset can then be prioritized for further characterization and evaluation for traits of interest such as commercial attributes or other traits of interest (van Zonneveld et al. 2012). This is similar to the *ex situ* concepts of establishing core collections (Frankel et al. 1995a; Odong et al. 2012). This could be combined with ecogeographic studies to assure the inclusion of populations that are adapted to different environments (Graudal et al. 1995; Parra-Quijano et al. 2011; Vinceti et al. 2013). In chapter 4, an ecogeographic analysis for wild potato species endemic to Bolivia was carried out.

Neutral markers are generally used to reveal patterns of gene flow and isolation between populations (Lowe et al. 2004). They provide insights in inbreeding rates and effective population sizes. These two parameters are correlated to population health. Indeed, significant correlations between molecular *alpha* diversity and fitness measures have been found in meta-analyses of studies that compare molecular diversity with quantitative variation (Reed and Frankham 2003; Vranckx et al. 2011). Also genetic differentiation between populations measured by molecular markers (Fst) and quantitative traits (Qst) is significantly correlated (Holderegger et al. 2006). However, the relation is far from straightforward. Processes detected by neutral and phenotypic indicators are being shaped by different factors and at different time scales. This explains why in a substantial number of natural tree species populations, molecular neutral diversity does not or even negatively correlate with diversity in adaptive traits (Holderegger et al. 2006).

During the domestication of cultivated plant species this phenomenon has been observed time and time again. During human selection, the variation in phenotypic expression on traits of interest increases, whereas the genetic base in populations under human selection is reduced due to genetic bottlenecks (Clement et al. 2010). Functional markers can help to identify variation in genome parts that are related to traits under human and natural selection. Enabling the application of these tools to a wide range of non-model plant species is expected to yield a major breakthrough in genetic diversity status and monitoring studies of plant species (Nichols and Neale 2010; Vinceti et al. 2013).

There remains an increasing need to learn more about phenotypic variation in adaptive characteristics and other functional traits of plant species, especially to understand the response of plant species to climate change (Hansen et al. 2012). Multi-site trials repeated over a number of consecutive growth cycles with crop varieties allow a cross-comparison of how different environmental and climatic conditions would affect the performance of specific accessions that are being conserved *ex situ*. The repeated recording of performance data from multi-site trials gives consistency to the predictive power of productivity models. It further allows improved calibration of these models themselves, by providing a real-world test of the performance of crops or varieties under different environmental scenarios. Information about performance will be especially important for understanding how crops and trees can be expected to perform in specific areas under climate change (i.e. under warmer conditions in combination with wetter or drier conditions). Several studies on the impact of climate change on performance have been carried out based on crop field trial data (Lobell et al. 2011; Ortiz et al. 2008) and tree species provenance trials (Saénz-Romero et al. 2006; Leibing et al. 2013). Based on such experiments, promising germplasm adapted to specific environments can be identified.

Monitoring diversity

There is a general concern that current agricultural production systems lead to loss of *in situ* PGR. Moreover, human disturbance, in general, leads to worldwide genetic erosion of plant species cultivated and natural populations (van de Wouw et al. 2010b; Graudal et al. *in prep*). However, status and trends in intra-specific diversity remain poorly quantified and require sound and easy-to-operate descriptors to estimate the diversity and conservation status in combination with indicators of pressure, response and benefit (Graudal et al. *in prep.*).

As argued in chapter 2 and in this chapter of the discussion, molecular markers are becoming increasingly accessible to carry out monitoring, especially when different types of molecular markers are used that are related to different mechanisms such as (1) stochastic processes that involve isolation and gene flow; and (2) genetic responses to natural and human selection (Newton et al. 1999; Hansen et al. 2012). Several studies have been carried out to understand genetic dynamics of crop species over several decades by comparing *ex situ* collections with newly sampled materials at the historical collecting sites (Gomez et al. 2005; Jensen et al. 2012). In long-lived perennials such as tree species, genetic variation of adult individuals may reflect past diversity and conservation status. They do not, however, provide much information on current adaptation and actual human disturbance such as fragmentation of forest fragments (Lowe et al. 2005). Therefore, it is recommended to monitor diversity and fitness of progeny material compared to that of adult trees (Lowe et al. 2005).

Citizen science can be a promising tool in monitoring plant genetic resources at species and perhaps also at botanical variety and crop varieties level. However, the latter may be complicated because of the small morphological differences that distinguish many crop varieties that only a few trained experts can detect. Some evidence is provided that farmers could have high competency values to distinguish between varieties (Benz et al. 2007). Therefore, farmers could be an ideal group to provide feedback about variety occurrence and performance (van Etten 2011). Citizen science has been the basis in many countries for monitoring animal biodiversity such as birds and has also been used to monitor plant species (Silvertown 2009). Customized and standardized online field guides are a pre-requisite to enable participants to identify plant species.

This discipline remains to be tested to see its reliability to map and monitor tropical tree species occurrence, and to distinguish between crop varieties, which may be difficult for non-specialists. This is especially a challenge when varieties are not well-documented, which is the case of local cultivars in centres of crop domestication. Therefore, it may be worthwhile to tap into the knowledge of experts and professionals that are working or have worked with crop varieties or tropical tree species for years, to name just two botanical areas. Much of their knowledge may not have been documented systematically. Ideally these specialists are organized in societies or networks such as LAFORGEN (www.laforgen.org) but not necessarily. A number of 2.0 web tools provide opportunities for interactions with networks and with interested persons in general.

One of the principles of citizen science is to validate the provided data by experts (Silvertown 2009). But data provided by expert networks also require validation. In chapter 3 of this dissertation, a method was presented to evaluate and formalize expert knowledge on species distribution on the basis of the culture consensus theory developed by Romney et al. (1986). The use of such methods opens the way to include more widely expert knowledge in mapping species occurrence, and validation of observed and modelled distribution and monitoring activities.

Finally, it is key that the collected and validated information goes back to users (the persons who gave the feedback, and other relevant groups) to support their field activities such as agricultural activities and conservation of plant diversity. In this way, feedback loops can be created that allow a constant update of information for users who can then act accordingly on the basis of the updated information.

Geographically targeting interventions to enhance conservation and use of cultivated plant diversity

Development trajectories can differ between rural communities because of distinct cultural and socio-economic contexts, and differences in biophysical conditions. In all these different contexts a sound use and management of cultivated plant diversity is the basis for sustainable agricultural production. But the solutions and interventions to optimize use and conservation of crop diversity differ according to cultural, socio-economic and biophysical circumstances. For example the level of irrigation potential and closeness to markets influences farmers' choices of crops and varieties. To which extent cultivated plant diversity is used and maintained depends also on the availability of local cultivated plant diversity; access to germplasm (for example from a genebank); the amount of knowledge about germplasm and its valorisation; the level of organization in rural communities; and/or market potential to develop high-value products from special plant species and varieties, among other factors (Jarvis et al. 2011). These factors may also be related to geospatial variables. This would allow developing a geographic classification of areas that require different types of interventions to improve use and management of cultivated plant diversity.

At global scale, several factors that drive cultivated plant diversity can be identified. First, centres of crop origin such as in Mexico, the Andes, the Fertile Crescent in the Middle East and eastern China are likely to be hotspots of cultivated plant diversity at species and intra-specific level. In these areas, early agricultural activities in the Holocene were the basis for human settlement and expansion, eventual development of complex cultures and crop dispersal to other areas (Meyer et al. 2012). Long histories of farmers' selections in historically cultural-rich locations have led to an accumulation of crop diversity over thousands of years in these areas. How agriculture originated independently in different areas, the reasons why people started plant cultivation and how many areas can be considered centres of plant domestication remains open to debate (Mannion 1999; Fuller 2010). Results from new archaeological, genetic and botanical data show, for example, that the number of areas where plants were domesticated is much higher than previously thought (Fuller 2010). Detailed reflections on why humans started to practise agricultural activities are outside the scope of this thesis. In short, there are two main hypotheses: (1) Climatic change around 16,000 to 12,000 years ago after the last ice age was a major driver for hunter-gatherer societies to alter their ways of food acquirement; and (2) As human population increased during that period, cultural factors became more important in human bands and tribes and were determinant for the initiation of plant cultivation and domestication (Mannion 1999). In addition to the use of molecular genetics to confirm or detect areas of high cultivated plant diversity, there are several data sources that can be used to detect these areas such as archaeological plant remains and crop representations in cultural expressions such as ceramics.

Secondly, socio-economic factors likely play an important role in cultivated plant diversity management. Generally, it is assumed that especially smallholders maintain intra-specific and inter-specific plant diversity as a risk management strategy for food security and/or income. However, also among smallholders within communities there are many differences. A meta-analysis of crop management in traditional rural communities from different countries across the globe revealed a positive relationship between farm area size and *alpha* varietal diversity and an increase in varietal divergence (*beta* diversity) as farm sizes become smaller (Jarvis et al. 2008). *Alpha* diversity at landscape level may increase as a consequence of high varietal divergence between farms (Willemen et al. 2007).

Thirdly, mountainous areas are expected to harbour high levels of cultivated plant diversity. One reason is that mountainous areas have different microclimates at landscape level. It requires thus a wide range of species and varieties to optimize mountain agricultural systems at landscape level (Tapia 2000). Ecogeographic analysis of traditional maize systems in Mexico shows that mid-elevation communities can adapt fairly easy their production systems to climate change through seed exchange with farmers within a 10 km radius where a wide range of different micro-climate can be found (Bellon et al. 2011). In contrast, highland and lowland systems that have less local micro-climate diversity require seed material from geographically more distant locations. The latter would require active support from governmental and non-governmental organizations (NGO) (Bellon et al. 2011). Another reason why these areas retain high levels of crop diversity could be that high-input and/or large-scale agriculture is not suitable for these areas. Mountainous areas have a low irrigation potential because they are often not suitable for the installation of extensive irrigation infrastructures. Either are they easily accessible for the introduction of improved varieties and fertilizer. Instead farmers rely under these conditions on locally adapted varieties and risk management strategies such as crop and varietal differentiation. Without doubt, high-input agriculture with high yielding varieties has had a big positive impact on global food security since it was introduced in the sixties (Evenson and Gollin 2003). However, this type of agriculture had and still has severe negative environmental and social impacts (Evenson and Gollin 2003; Chhetri and Chaudhary 2011). In the areas where this type of agriculture has been introduced, it is also held responsible for substantial losses of agricultural biodiversity (van de Wouw et al. 2010a). These observations raise questions about the sustainability of this type of agriculture. Therefore there is a need to search for alternative more sustainable solutions to feed the world's human population (Godfray et al. 2010). A better use of cultivated plant diversity can be part of a multi-faceted strategy to accomplish that (Godfray et al. 2010).

Fourthly, biological factors influence intra-specific diversity of specific crops. Gene flow between cultivated plants and their wild progenitors or relatives in overlapping areas of distribution causes elevated levels of intra-specific cultivated plant diversity. In the Fertile Crescent, for example, higher levels of molecular diversity of domesticated emmer wheat (*Triticum turgidum* L. subsp. *dicoccon* (Schrank) Thell.) and bread wheat (*T. aestivum* L.) have been found outside their centres of domestication (Dvorak et al. 2011). Introgression between the domesticates and their wild ancestors in sympatric areas of distribution explains these increased levels of crop genetic diversity (Dvorak et al. 2011).

The above-mentioned factors are generally recognized to drive cultivated plant diversity. Probably there are additional agencies as well that influence dynamics in cultivated plant diversity. However, little work has been done on developing spatial models that try to explain the role of different variables in shaping geographic patterns of cultivated plant diversity at national, regional or global level. To develop these models, sound georeferenced datasets of inter-specific and intra-specific with minimum sampling bias are required. Hence, it is important to share data, standardize and use sound and easy-to-implement descriptors of plant diversity.

Another challenge is to determine the biophysical and socio-economic characteristics of agricultural systems that can benefit from crop and varietal diversification and to understand the requirements of an enabling environment that empowers farmers in making information-based decisions to adopt new varieties and crops. Recently, a heuristic framework to identify actions to enhance use and conservation of crop diversity in different types of agricultural production systems has been developed (Jarvis et al. 2011). It will be interesting to develop spatial models on the basis of such frameworks. This will help to identify interventions for rural areas with different cultural, socio-economic and biophysical characteristics.

At landscape and farmer level, seed exchange between smallholders is an important factor in the dynamics and use of cultivated plant diversity (Thomas et al. 2011). In chapter 2 of this dissertation on the application of molecular marker characterization for PGR *in situ* conservation, differences in seed systems and local geographic distributions of intra-specific diversity were observed between and within communities and for different crops (e.g. Barry et al. 2007; de Haan et al. 2009a; Worthington et al. 2012). So although it would be possible to detect general patterns in conservation and use of crop genetic diversity at global, regional or national level, each local situation has its unique characteristics (Jarvis et al. 2011). Successful implementation of interventions will therefore require involvement of local governmental bodies, farmer associations and NGOs (Jarvis et al. 2011).

Funding

Without the financial support of several institutions and research programs the studies that are presented in this dissertation could not have been accomplished.

The CGIAR Research Programme on Forests, Trees and Agroforestry supported writing and data analysis of chapter 2 ´Application of molecular marker technology in spatial analysis to optimize in situ conservation of plant genetic resources´; chapter 3 ´Application of consensus theory to formalize expert evaluations of plant distribution models for better species´ conservation´; chapter 5 ´Diversity, distribution, germplasm conservation and use of peach palm (*Bactris gasipaes* Kunth); and chapter 6 ´Mapping genetic diversity of cherimoya (*Annona cherimola* Mill.): application of spatial analysis for conservation and use of plant genetic resources.

The project ´Strengthening Regional Collaboration in Conservation and Sustainable Use of Forest Genetic Resources in Latin America and Sub-Saharan Africa´ financed by INIA Spain supported research design, data analysis and writing of chapter 3 and 6.

NUFFIC and the United Nations Environment Programme (UNEP)/Global Environmental Facility (GEF) -supported project *´In situ* conservation of wild crop relatives through enhanced information management and field application´ supported field work for chapter 4 ´Endemic wild potato (*Solanum* spp.) biodiversity status in Bolivia: reasons for conservation concerns´.

The Amazon Initiative, the Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) and INIA Spain supported writing and data analysis of chapter 5.

The sixth framework research program of the European Commission supported within the context of international co-operation activities (INCO), research design, field work and molecular analysis of the samples for chapter 6.

Bibliography

Adin A, Weber JC, Sotelo Montes C, Vidaurre H, Vosman B, et al. (2004) Genetic differentiation and trade among populations of peach palm (*Bactris gasipaes* Kunth) in the Peruvian Amazon - implications for genetic resource management. Theoretical Applied Genetics 108: 1564-1573.

Alercia A, Diulgheroff S, Mackay M (2012) FAO/Bioversity multi-crop passport descriptors V.2 [MCPD V.2]. Bioversity International, Rome.

Altieri MA, Merrick LC (1987) *In situ* conservation of crop genetic resources through maintenance of traditional farming. Economic Botany 41: 86-96.

Alves-Pereira A, Clement CR, Picanco-Rodrigues D (2012) Genetic divergence among populations and accessions of the spineless peach palm from Pampa Hermosa landrace used in the heart-of-palm agribusiness in Brazil. Genetics and Molecular Biology 35: 474–479.

Anderson RP, Lew D, Peterson AT (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling 162: 211-232.

Ansley RJ, Boutton TW, Mirik M, Castellano MJ, Kramp BA, et al. (2010) Restoration of C4 grasses with seasonal fires in a C3/C4 grassland invaded by *Prosopis glandulosa*, a fire-resistant shrub. Applied Vegetation Science 13: 520-530.

APG III (2009) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161: 105-121.

Araújo MB, Peterson AT (2012) Uses and misuses of bioclimatic envelope modelling. Ecology 93: 1527-1539.

Araújo MB, Pearson RG, Thuiller W, Erhard M (2005) Validation of species-climate impact models under climate change. Global Change Biology 11: 1504–1513.

Araújo de CM, Rodrigues DP, Filho SA, Clement CR (2010) Genetic variability in the peach palm genebank with RAPD markers. Crop Breeding and Applied Biotechnology 10: 211-217.

Atlas Florae Europaeae (2011) Distribution of Vascular Plants in Europe. Helsinki: Finnish Museum of Natural History, University of Helsinki.

Austin M (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling 200: 1-19.

Avise JC (2010) Perspective: conservation genetics enters the genomics era. Conservation Genetics 11: 665-669.

Azpilicueta MM, Gallo LA, van Zonneveld M, Thomas E, Moreno C, et al. (2013) Management of *Nothofagus* genetic resources: definition of genetic zones based on a combination of nuclear and chloroplast marker data. Forest Ecology and Management *Accepted*.

Baillie JEM, Hilton-Taylor C, Stuart SN (2004) IUCN Red List of Threatened Species: A Global Species Assessment. IUCN, Gland, Switzerland and Cambridge, UK.

Banks WE, Aubry T, d'Errico F, Zilhão J, Lira-Noriega A, et al. (2011) Eco-cultural niches of the Badegoulian: Unraveling links between cultural adaptation and ecology during the Last Glacial Maximum in France. Journal of Anthropological Archaeology 30: 359–374.

Barry S, Elith J (2006) Error and uncertainty in habitat models. Journal of Applied Ecology 43: 413–423.

Barry MB, Pham JL, Courtois B, Billot C, Ahmadi N (2007) Rice genetic diversity at farm and village levels and genetic structure of local varieties reveal need for *in situ* conservation. Genetic Resources and Crop Evolution 54: 1675-1690.

Barthlott W, Hostert A, Kier G, Küper W, Kreft H, et al. (2007) Geographic patterns of vascular plant diversity at continental to global scales. Erdkunde 61: 305-317.

Batjes NH (2005) SOTER-based soil parameter estimates for Latin America and the Caribbean (ver. 1.0). Report 2005/02. ISRIC - World Soil Information, Wageningen.

Beauvais GP, Keinath DA, Hernandez P, Master L, Thurston R, et al. (2006) Element Distribution Modeling: A Primer (Version 2). Natureserve, Arlington, Virginia.

Bellon MR, Hodson D, Hellin J (2011) Assessing the vulnerability of traditional maize seed systems in Mexico to climate change. Proceedings of the National Academy of Sciences 108: 13432-13437.

Bhullar NK, Street K, Mackay M, Yahiaouia N, Kellera B (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. Proceedings of the National Academy of Sciences 106: 9519–9524.

Bierman SM, Butler A, Marion G, Kühn I (2010) Bayesian image restoration models for combining expert knowledge on recording activity with species distribution data. Ecography 33: 451-460.

Bioversity International (2009) Modern crop varieties can increase local genetic diversity. Plant Breeding News, an Electronic Newsletter of Applied Plant Breeding, edition 201, May 2009.

Bioversity International (2004) New World Fruit Database. <http://www.bioversityinternational.org/databases/new_world_fruits_database> Accessed 16 January 2013.

Blach-Overgaard A, Svenning JC, Dransfield J, Greve M, Balslev H (2010) Determinants of palm species distributions across Africa: the relative roles of climate, non-climate environmental factors, and spatial constraints. Ecography 33: 380–391.

Bonavia D, Ochoa CM, Tovar SÓ, Palomino RC (2004) Archaeological evidence of cherimoya (*Annona cherimolia* Mill.) and guanabana (*Annona muricata* L.) in ancient Peru. Economic Botany 58: 509-522.

Boshier DH, Gordon JE, Barrance AJ (2004) Prospects for *circa situm* tree conservation in Mesoamerican dry-forest agro-ecosystems. In: Frankie GW, Mata A, Vinson SB, editors. Biodiversity conservation in Costa Rica. University of California Press. Berkeley , California. 210–226.

Botero Botero L, Atehortua L (1999) Propagación in vitro de la palma de chontaduro (*Bactris gasipaes* H.B.K) por embriogénesis somática. Memorias del VI Seminario Nacional y II Internacional de Recursos Vegetales Promisorios. Universidad del Chocó, Quibdó. 95–100.

Bradley BA, Marvin DC (2011) Using expert knowledge to satisfy data needs: mapping invasive plant distributions in The Western United States. Western North American Naturalist 71: 302-315.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and Regression Trees. Chapman and Hall, New York.

Breiman L (2001) Random forests. Machine Learning 45: 5–32.

Bremer B, Bremer K, Chase MW, Fay M, Reveal J (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161: 105-121.

Briggs D, Walter M (1999) Evolution: some general considerations. In: Plant variation and evolution. 3rd edition, Cambridge University Press, UK. 367-398.

Benz B, Perales H, Brush S (2007) Tzeltal and tzotzil farmer knowledge and maize diversity in Chiapas, Mexico. Current Anthropology 48: 289-300.

Bowe C, Haq N (2010) Quantifying the global environmental niche of an underutilised tropical fruit tree (*Tamarindus indica*) using herbarium records. Agriculture, Ecosystems and Environment 139: 51–58.

Brown AHD, Hodgkin T (2008) Measuring, managing and maintaining crop genetic diversity on farm. In: Jarvis DI, Padoch C, Cooper HD, editors. Managing Biodiversity in Agricultural Ecosystems. Columbia University Press. 13-33.

Brummitt N, Bachman S, Moat J (2008) Applications of the IUCN Red List: towards a global barometer for plant diversity. Endangered Species Research 6: 127-135.

Busby JR (1991) BIOCLIM – a bioclimate analysis and prediction system. In: Margules CR, Austin MP, editors. Nature conservation: cost effective biological surveys and data analysis. CSIRO, Melbourne. 64-68.

Bydekerke L, Van Ranst E, Scheldeman X, Van Damme P (1999) Edaphoclimatological study of cherimoya (*Annona cherimola* Mill.) in Loja province, Southern Ecuador. Acta Horticulturae 497: 71-90.

Byers DL (2005) Evolution in heterogeneous environments and the potential of maintenance of genetic variation in traits of adaptive significance. Georgia Genetics Review III 3: 107–124.

Cadima X, Gandarillas A (2009) Conserving Bolivia's potatoes. In: Geneflow Bioversity International. p 20.

Camadro EL (2011) Hibridación y flujo génico en especies silvestres de papa de la Argentina. Journal of Basic and Applied Genetics 22: 1-5.

Camadro EL, Erazzú LE, Maune JF, Bedogni MC (2012) A genetic approach to the species problem in wild potato. Plant Biology 14: 543-554.

Canhos VP, Souza S, Giovanni R, Canhos DAL (2004) Global biodiversity informatics: setting the scene for a "new world" of ecological modeling. Biodiversity Informatics 1: 1–13.

Carpenter G, Gillison AN, Winter J (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation 2: 667–680.

CBD (1992) Convention on Biological Diversity. <http://www.cbd.int/convention/text>. Accessed on April 17th 2013.

CBD (2010) Aichi Biodiversity Targets. < https://www.cbd.int/sp/targets>. Accessed at April 17th 2013.

CBD (2012) Updated Global Strategy for Plant Conservation 2011-2020. <https://www.cbd.int/gspc>. Accessed at April 17th 2013.

Chan LM, Brown JL, Yoder AD (2011) Integrating statistical genetic and geospatial methods brings new power to phylogeography. Molecular Phylogenetics and Evolution 59: 523-537.

Chapman AD (2005) Principles and methods of data cleaning - primary species and species-occurrence data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

CHERLA (2008) Inventory of current *ex situ* germplasm collections. Deliverable 7, Project no. 015100, INCO sixth framework programme.

Chhetri N, Chaudhary P (2011) Green revolution: pathways to food security in an era of climate variability and change? Journal of Disaster Research 6: 486-497.

Clement CR, Santos RP, Desmouliere SJM, Ferreira EJL, Farias Neto JT, et al. (2009) Ecological adaptation of wild peach palm, its *in situ* conservation and deforestation-mediated extinction in southern Brazilian Amazonia. PLOS ONE 4: e4564. Clement (1999) 1942 and the loss of Amazonian crop genetic resources. I. the relation between domestication and human population decline. Economic Botany 53: 188-202.

Clement CR, Yuyama K, Chávez Flores WB (2001) Recursos genéticos de pupunha (Genetic resources of pejibaye). In: Sousa NR, Souza AGC, editors. Recursos fitogenéticos na Amazônia Ocidental: conservação, pesquisa e utilização. Embrapa Amazônia Ocidental, Manaus. 143-187.

Clement CR, Weber JC, van Leeuwen J, Astorga Domian C, Cole DM, et al. (2004) Why extensive research and development did not promote use of peach palm fruit in Latin America. Agroforestry Systems 61: 195-206.

Clement CR, de Cristo-Araújo M, Coppens d'Eeckenbrugge G, Alves Pereira A, Picanço D, et al. (2010) Origin and domestication of native Amazonian crops. Diversity 2: 73-106.

Cleveland DA, Soleri D (2007) Extending Darwin's analogy: bridging differences in concepts of selection between farmers, biologists, and plant breeders. Economic Botany 61: 121-136.

Cobben MMP, Verboom J, Opdam PFM, Hoekstra RF, Jochem R, et al. (2011) Projected climate change causes loss and redistribution of genetic diversity in a model metapopulation of a medium-good disperser. Ecography 34: 920-932.

Cole DM, White TL, Nair PKR (2007) Maintaining genetic resources of peach palm (*Bactris gasipaes* Kunth): The role of seed migration and swidden-fallow management in northeastern Peru. Resources and Crop Evolution 54: 189–204.

Coleman WK (2008) Evaluation of wild *Solanum* species for drought resistance 1. *Solanum gandarillasii* Cardenas. Environmental and Experimental Botany 62: 221-230.

Cornelius JP, Clement CR, Weber JC, Sotelo-Montes C, van Leeuwen J, et al. (2006) The trade-off between genetic gain and conservation in a participatory improvement programme: the case of peach palm (*Bactris gasipaes* Kunth). Forests, Trees and Livelihoods 16: 17–34.

Cornelius JP, Weber JC, Sotelo-Montes C, Ugarte-Guerra LJ (2010) Phenotypic correlations and site effects in a Peruvian landrace of peach palm (*Bactris gasipaes* Kunth). Euphytica 173: 173–183.

Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20: 273–297.

Coudun C, Gégout JC, Piedallu C, Rameau JC (2006) Soil nutritional factors improve models of plant species distribution: an illustration with *Acer campestre* (L.) in France. Journal of Biogeography 33: 1750-1763.

Couvreur TLP, Bilotte N, Risterucci AM, Lara C, Vigouroux Y, et al. (2006) Close genetic proximity between cultivated and wild *Bactris gasipaes* Kunth revealed by microsatellite markers in Western Ecuador. Genetic Resources and Crop Evolution 53: 1361–1373.

Couvreur TLP, Hahn WJ, de Granville J-J, Pahm J-L, Ludeña B, et al. (2007) Phylogenetic relationships of the cultivated neotropical palm *Bactris gasipaes* (Arecaceae) with its wild relatives inferred from chloroplast and nuclear DNA polymorphisms. Systematic Botany 32: 519-530.

Da Silva JBF, Clement CR (2005) Wild pejibaye (*Bactris gasipaes* Kunth var. *chichagui*) in Southeastern Amazonia. Acta Botanica Brasilica 19: 281–284.

Dawson IK, Lengkeek A, Weber JC, Jamnadass R (2009) Managing genetic variation in tropical trees: linking knowledge with action in agroforestry ecosystems for improved conservation and enhanced livelihoods. Biodiversty and Conservation 18: 969-986.

Dawson IK, Vinceti B, Weber JC, Neufeldt H, Russell J, et al. (2011) Climate change and tree genetic resource management: maintaining and enhancing the productivity and value of smallholder tropical agroforestry landscapes: A review. Agroforest Systems 81: 67-78.

Dawson IK, Guariguata MR, Loo J, Weber JC, Lengkeek A, et al. (2013) What is the relevance of smallholders' agroforestry systems for conserving tropical tree species and genetic diversity in *circa situm*, *in situ* and *ex situ* settings? A review. Biodiversity Conservation 22: 301–324.

de Haan S, Núñez J, Bonierbale M, Ghislain M (2009a) Species, morphological and molecular diversity of Andean potatoes in Huancavelica, central Peru. In: Potato diversity at height: multiple dimensions of farmer-driven in-situ conservation in the Andes. PhD thesis, Wageningen University, The Netherlands. 35-58.

de Haan S, Bonierbale M, Juárez H, Poma J, Salas E, et al. (2009b) Annual spatial management of potato diversity in Peru's central Andes. In: Potato diversity at height: multiple dimensions of farmer-driven in-situ conservation in the Andes. PhD thesis, Wageningen University, The Netherlands. 91-115.

De Vicente MC, Fulton T (2004) Using Molecular Marker Technology in Studies on Plant Genetic Diversity: Learning module Volume 1. International Plant Genetic Resources Institute, Rome and Cornell University, Ithaca, New York.

De Vicente MC, Lopez C, Fulton T (2004b) Genetic Diversity Analysis with Molecular Marker Data: Learning Module Volume 2. International Plant Genetic Resources Institute, Rome and Cornell University, Ithaca, New York.

Degen B, Scholz F (1998) Spatial genetic differentiation among populations of European beech (*Fagus sylvatica* L.) in western Germany as identified by geostatistical analysis. Forest Genetics 5: 191–199.

Del Rio AH, Bamberg JB, Huaman Z, Salas A, Vega SE, et al. (2001) Association of ecogeographical variables and RAPD marker variation in wild potato populations of the USA. Crop Science 41: 870-878.

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. Nature 418: 700-707.

Dobson AJ (2002) An introduction to generalized linear models. 2nd Edition. Chapman and Hall/CRC, London.

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, et al. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30: 609-628.

Dulloo ME, Hunter D, Borelli T (2010) *Ex situ* and *in situ* conservation of agricultural biodiversity: Major advances and research needs. Notulae Botanicae Horti Agrobotanici Cluj-Napoca 38: 123-135.

Dvorak J, Luo, M-Ch, Akhunov ED (2011) N.I. Vavilov’s theory of centres of diversity in the light of current understanding of wheat diversity, domestication and evolution. Czech Journal of Genetics and Plant Breeding 47: S20–S27.

Eaton D, Windig J, Hiemstra SJ, van Veller M, Trach NX, et al. (2006) Indicators for Livestock and Crop Biodiversity. CGN report 2006/05. Centre for Genetic Resources, CGN/DLO Foundation, Wageningen, The Netherlands.

Eding H, Crooijmans RPMA, Groenen MAM, Meuwissen THE (2002) Assessing the contribution of breeds to genetic diversity in conservation schemes. Genetics Selection, Evolution 34: 613-633.

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, et al. (2006) Novel methods improve prediction of species’ distributions from occurrence data. Ecography 29: 129–151.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, et al. (2011) A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17: 43-57.

Ellstrand NC, Heredia SM, Leak-Garcia JA, Heraty JM, Burger JC, et al. (2010) Crops gone wild: evolution of weeds and invasives from domesticated ancestors. Evolutionary Applications 2010: 494-504.

Eshbaugh WH (2012) The taxonomy of the genus *Capsicum*. In: Russo VM, editor. Peppers, Production and Uses. CABI. 1-13.

Escribano P, Viruel MA, Hormaza JI (2004) Characterization and cross-species amplification of microsatellite markers in cherimoya (*Annona cherimola* Mill. Annonaceae). Molecular Ecology Notes 4: 746–748.

Escribano P, Viruel MA, Hormaza JI (2007) Molecular analysis of genetic diversity and geographic origin within an *ex situ* germplasm collection of cherimoya by using SSRs. Journal of the American Society for Horticulture Science 132: 357-367.

Escribano P, Viruel MA, Hormaza JI (2008a) PERMANENT GENETIC RESOURCES: Development of 52 new polymorphic SSR markers from cherimoya (*Annona cherimola* Mill.). Transferability to related taxa and selection of a reduced set for DNA fingerprinting and diversity studies. Molecular Ecology Resources 8: 317–321.

Escribano P, Viruel MA, Hormaza JI (2008b) Comparison of different methods to construct a core germplasm collection in woody perennial species with simple sequence repeat markers. A case study in cherimoya (*Annona cherimola*, Annonaceae), an underutilised subtropical fruit tree species. Annals of Applied Biology 153: 25–32.

Escudero A, Iriondo JM, Torres ME (2003) Spatial analysis of genetic diversity as a tool for plant conservation. Biological Conservation 113: 351-365.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611–2620.

Evenson RE, Gollin D (2003) Assessing the impact of the green revolution, 1960 to 2000. Science 300: 758-762.

Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. Nature Reviews Genetics 7: 745–758.

FAO (1988) Soil map of the world. Revised legend. Reprinted with corrections. World Soil Resources Report 60. FAO, Rome.

FAO (2007) Ecocrop. <www.ecocrop.fao.org>. Accessed at January 5th 2013.

FAO (2009) The International Treaty on Plant Genetic Resources for Food and Agriculture. <http://www.planttreaty.org/content/texts-treaty-official-versions> Accessed at April 17th 2013.

FAO (2010a) The second report on the state of the world's plant genetic resources for food and agriculture. Rome. < http://www.fao.org/agriculture/seed/sow2/en/>. Accessed at April 17th 2013.

FAO (2010b) Global forest resources assessment 2010. Rome. <http://www.fao.org/forestry/fra/fra2010/en/>. Accessed at April 17th 2013.

FAO (2011) Draft updated global plan of action for the conservation and sustainable utilization of plant genetic resources for food and agriculture. Fifth session of the Intergovernmental Technical Working Group on Plant Genetic Resources for Food and Agriculture, Rome, 27-29 April 2011.

Farré JM, Hermoso JM, Guirado E, García-Tapia J (1999) Techniques of cherimoya cultivation in Spain. Acta Horticulturae 497: 91-118.

Ferreira E (1999) The phylogeny of pupunha (*Bactris gasipaes* Kunth, Palmae) and allied species. In: Henderson A, Borchsenius F, editors. Evolution, Variation and Classification of palms. Volume 83. Memoirs of the New York Botanical Garden, New York. 225–236.

Feely KJ, Silman MR (2011) Keep collecting: accurate species distribution modelling requires more collections than previously thought. Diversity and Distributions 2011: 1–9.

Field R, Hawkins BA, Cornell HV, Currie DJ, Diniz-Filho JAF, et al. (2008) Spatial species-richness gradients across scales: a meta-analysis. Journal of Biogeography 36: 132-147.

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24: 38-49.

Finkeldey R (2005) An Introduction to Tropical Forest Genetics. Institute of Forest Genetics and Forest Tree Breeding. Georg-August-University Gottingen, Gottingen, Germany.

FRA (Forest Resources Assessment Programme) (2001) Global ecological zoning for the global forest resources assessment 2000. Final Report. Working Paper 56. FAO. Rome.

Franco J, Crossa J, Desphande S (2010) Hierarchical multiple-factor analysis for classifying genotypes based on phenotypic and genetic data. Crop Science 50: 105–117.

Frankel OH, Brown AHD, Burdon J (1995a) The conservation of cultivated plants. In: Frankel OH, Brown AHD, Burdon JJ, editors. The Conservation of Plant Biodiversity. 1st edition. University Press, Cambridge, UK. 79–117.

Frankel OH, Brown AHD, Burdon J (1995b) The genetic diversity of wild plants. In: Frankel OH, Brown AHD, Burdon JJ, editors. The Conservation of Plant Biodiversity. 1st edition. University Press, Cambridge, UK. 10-38.

Frankel OH, Brown AHD, Burdon J (1995c) Plant species conservation and population biology. In: Frankel OH, Brown AHD, Burdon JJ, editors. The Conservation of Plant Biodiversity. 1st edition. University Press, Cambridge, UK. 118-147.

Franklin SE, Wilson BA (1991) Vegetation mapping and change detection using SPOT MLA and LandSat imagery in Kluane National Park. Canadian Journal of Remote Sensing 17: 2–22.

Fritz S, Bartholomé E, Belward A, Hartley A, Stibig HJ, et al. (2003) Harmonisation, mosaicing and production of the Global Land Cover 2000 database (Beta Version). EUR 20849 EN. European Commission, Joint Research Centre.

Fuller DQ (2010) An emerging paradigm shift in the origins of agriculture. General Anthropology 17: 1-12.

Gallo LA, Marchelli P, Chauchard L, Gonzalez Peñalba M (2009) Knowing and doing: research leading to action in the conservation of forest genetic diversity of Patagonian temperate forests. Conservation Biology 23: 895–898.

GBIF (2012) Available: www.gbif.org .

GCDT (Global Crop Diversity Trust) (2010) Anual Report 2010. Rome.

Geburek T, Turok J, editors (2005) Conservation and management of forest genetic resources in Europe. Arbora Publishers, Zvolen, Slovakia.

Gepts P (2003) Crop domestication as a long-term selection experiment. In: Janick J, editor. Plant breeding reviews 24 Part 2: Long-term selection: crops, animals, and bacteria. 1-44.

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D., et al. (2010). Food security: the challenge of feeding 9 billion people. Science 327: 812-818.

Gómez OJ, Blair MW, Frankow-Lindberg BE, Gullberg U (2005) Comparative study of common bean (*Phaseolus vulgaris* L.) landraces conserved *ex situ* in genebanks and *in situ* by farmers. Genetic Resources and Crop Evolution 52: 371–380.

González-Orozco CE, Brown AHD, Knerr N, Miller JT, Doyle JJ (2012). Hotspots of diversity of wild Australian soybean relatives and their conservation *in situ*. Conservation Genetics 13: 1269–1281.

Gorenflo LJ, Romaine S, Mittermeier RA, Walker-Painemill (2012) Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas. Proceedings of the National Academy of Science 109: 8032–8037.

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4: 379–391.

Gotelli NJ, Entsminger GL (2004) EcoSim: Null Modeling Software for Ecologists. Version 7. Acquired Intelligence Inc. Available: <http://garyentsminger.com/ecosim/index.htm> . Accessed April 17th 2013.

Graefe S, Dufour D, van Zonneveld M, Rodriguez F, Gonzalez A (2013) Peach palm (*Bactris gasipaes*) in tropical Latin America: implications for biodiversity conservation, natural resource management and human nutrition. Biodiversity and Conservation 22: 269-300.

Graham CH, Hijmans RJ (2006) A comparison of methods for mapping species ranges and species richness. Global Ecology and Biogeography 15: 578-587.

Graudal L, Kjær ED, Canger S (1995) A systematic approach to the conservation of genetic resources of trees and shrubs in Denmark. Forest Ecology and Management 73: 117–134.

Graudal L, Yanchuk A, Kjær E (2004) National planning. In: FAO, FLD, IPGRI, editors. Forest genetic resources conservation and management. Volume 1: Overview, concepts and some systematic approaches. International Plant Genetic Resources Institute, Rome, Italy. 25-36.

Grum M, Atieno G (2007) Statistical Analysis for Plant Genetic Resources: Clustering and Indices in R Made Simple. Handbooks for Genebanks, No. 9. Bioversity International, Rome.

Guarino L (1995) Mapping the ecogeographic distribution of biodiversity. In: Guarino L, Ramanatha Rao V, Reid R, editors. Collecting Plant Diversity: Technical Guidelines. CABI. 287-314.

Guarino L, Jarvis A, Hijmans RJ, Maxted N (2002) Geographic information systems (GIS) and the conservation and use of plant genetic resources. In: Engels JMM, Ramanatha Rao V, Brown AHD, Jackson MT, editors, Managing plant genetic diversity. International Plant Genetic Resources Institute (IPGRI), Rome 387-404.

Guarino L, Maxted N, Chiwona EA (2005) A Methodological Model for Ecogeographic Surveys of Crops. IPGRI Technical Bulletin No. 9. International Plant Genetic Resources Institute (IPGRI), Rome.

Guisan A, Zimmerman NE (2000) Predictive habitat distribution models in ecology. Ecological Modelling 135: 147–186.

Guisan, A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993-1009.

Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: A survey of developments over the last 20 years. Euphytica 156: 1-13.

Hajjar R, Jarvis DI, Gemmill-Herren B (2008) The utility of crop genetic diversity in maintaining ecosystem services. Agriculture, Ecosystems and Environment 123: 261–270.

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4: 4–9.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. Philosophical Transactions of the Royal Society of London B 351: 1291-1298.

Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, et al. (2002) African pastoralism: genetic imprints of origins and migrations. Science 296: 336–339.

Hansen, MM, Olivieri I, Waller DM (2012) Monitoring adaptive genetic responses to environmental change. Molecular Ecology 21: 1311-1329.

Hastie T, Tibshirani R (1986) Generalized additive models. Statistical Science 1: 297–318.

Hauptvogel R, Kuna R, Štrba P, Hauptvogel P (2010) GIS design for *in situ* conservation of rare and endangered species. Czech Journal of Genetics and Plant Breeding 46: 50-53.

Hawkes JG, Hjerting JP (1989) The potatoes of Bolivia: Their breeding and evolutionary relationships. Oxford, UK, Oxford University Press.

Hawkes JG (1990) The potato: evolution, biodiversity and genetic resources. London: Press Belhaven.

Henderson A (2000) Bactris (Palmae). Flora Neotropica 79.New York Botanical Garden, US.

Hernandez PA, Graham CH, Master LL, Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29: 773–785.

Hernández-Ugalde JA, Mora Urpí J, Rocha Nuñez O (2008) Diversidad genética y relaciones de parentesco de las poblaciones silvestres y cultivadas de pejibaye (*Bactris gasipaes*, Palmae), utilizando marcadores microsatelites. Revista Biologica Tropical 56: 217-245.

Hernández-Ugalde JA, Mora Urpí J, Rocha Nuñez O (2011) Genetic relationships among wild and cultivated populations of peach palm (*Bactris gasipaes* Kunth, Palmae): evidence for multiple independent domestication events. Genetic Resources and Crop Evolution 58: 571-583.

Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93: 679-688.

Hijmans RJ, Garrett KA, Huamán Z, Zhang DP, Schreuder M, et al. (2000) Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. Conservation Biology 14: 1755–1765.

Hijmans RJ, Spooner DM (2001) Geographic distribution of wild potato species. American Journal of Botany 88: 2101-2112.

Hijmans RJ, Spooner DM, Salas AR, Guarino L, de la Cruz J, et al. (2002) Atlas of wild potatoes: Systematic and ecogeographic studies on crop genepools. Rome, Italy: International Plant Genetic Resources Institute.

Hijmans RJ, Guarino L, Jarvis A, O'Brien R, Mathur P, et al. (2005a) DIVA-GIS version 5.2 manual.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, et al. (2005b) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965–1978.

Hijmans RJ, Gavrilenko T, Stephenson S, Bamberg J, Salas A, et al. (2007) Geographical and environmental range expansion through polyploidy in wild potatoes (*Solanum* section *Petota*). Global Ecology and Biogeography 16: 485–495.

Hijmans RJ, Phillips S, Leathwick J, Elith J (2012) Dismo: Species distribution modeling. R package version 0.7-17. <http://CRAN.R-project.org/package=dismo>. Accessed at April 17th.

Hijmans RJ, van Etten J (2012) Geographic analysis and modeling with raster data. R package version 2.1-25. <http://cran.r-project.org/web/packages/raster/index.html>. Accessed at April 17th.

Hirota M, Holmgren M, van Nes EH, Scheffer M (2011) Global resilience of tropical forest and savanna to critical transitions. Science 334: 232-235.

Hoffmann MH, Glaß AS, Tomiuk J, Schmuths H, Fritsch RM, et al. (2003) Analysis of molecular data of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS). Molecular Ecology 12: 1007–1019.

Hoffmann M, Hilton-Taylor C, Angulo A (2010) The impact of conservation on the status of the world's vertebrates. Science 330: 1503-1509.

Hoffmann AA, Sgró CN (2011) Climate change and evolutionary adaptation. Nature 479: 479-485.

Höft M, Barik SK, Lykke AM (1999) Quantitative Ethnobotany: Applications of Multivariate and Statistical Analyses in Ethnobotany. People and Plants Working Paper No. 6. United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris.

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. Landscape Ecology: 797–807.

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. Trends in Plant Science 15: 675–683.

Holland SM (2008) Non-metric multidimensional scaling (MDS). Department of Geology, University of Georgia, Athens, Georgia.

Hollingsworth PM, Dawson IK, Goodall-Copestake WP (2005) Do farmers reduce genetic diversity when they domesticate tropical trees? A case study from Amazonia. Molecular Ecology 14: 497-501.

Hurlbert AH, Jetz W (2007) Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. Proceedings of the National Academy of Science 104: 13384-13389.

Hunziker AT (2001) The genera of Solanaceae illustrated, arranged according to a new system. Gantner Verlag, Ruggel, Liechtenstein.

Ibisch PL, Mérida G, editors (2003) Biodiversidad: La riqueza de Bolivia. Estado de conocimiento y conservación. Santa Cruz de la Sierra, Bolivia. Ministerio de Desarrollo Sostenible. Editorial FAN.

Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. New Phytologist 189: 909-922.

Işik, K. (2011) Rare and endemic species: Why are they prone to extinction? Turkish Journal of Botany 35: 411-417.

IUCN (2010) Guidelines for Using the IUCN Red List Categories and Criteria, Version 8.1. Standards and Petitions Subcommittee of the International Union for Conservation of Nature (IUCN) Species Survival Commission.

IUCN (2012) IUCN Red List of Threatened Species. Version 2012.2. <www.iucnredlist.org>. Accessed at November 27th 2012.

Jansky SH, Simon R, Spooner DM (2008) A test of taxonomic predictivity: Resistance to early blight in wild relatives of cultivated potato. Phytopathology 98: 680-687.

Jarvis A, Ferguson ME, Williams DE, Guarino L, Jones PG, et al. (2003) Biogeography of wild Arachis: assessing conservation status and setting future priorities. Crop Science 43: 1100–1108.

Jarvis A, Williams K, Williams BD, Guarino L, Caballero PJ, et al. (2005) Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. Genetic Resources and Crop Evolution 52: 671–682.

Jarvis A, Lane A, Hijmans RJ (2008) The effect of climate change on crop wild relatives. Agriculture, Ecosystems and Environment 126: 13–23.

Jarvis A, Touval JL, Castro Schmitz M, Sotomayor L, Hyman GG, et al. (2010) Assessment of threats to ecosystems in South America. Journal for Nature Conservation 18: 180–188.

Jarvis DI, Brown AHD, Cuong PH, Collado-Panduro L, Latournerie-Moreno L, et al. (2008) A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. Proceedings of the National Academy of Sciences 105: 5326-533.

Jarvis D, Hodgkin T, Sthapit BR, Fadda C, Lopez-Noriega I (2011) An heuristic framework for identifying multiple ways of supporting the conservation and use of traditional crop varieties within the agricultural production system. Critical Reviews in Plant Sciences 30: 125-176.

Jensen HR, Dreiseitl A, Sadiki M, Schoen DJ (2012) The Red Queen and the seed bank: pathogen resistance of ex situ and in situ conserved barley. Evolutionary Applications 5: 353-367.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403-1405.

Jombart, T (2012) adegenet: an R package for the exploratory analysis of genetic and genomic data. R package 'Adegenet' version 1.3-7 <http://cran.r-project.org/web/packages/adegenet/>. Accessed at April 17th 2013.

Jongman RHG, Ter Braak CJF, Van Tongeren OFR (1995) Data Analysis in Community and Landscape Ecology. Cambridge University Press, Cambridge, UK.

Jones PG, Guarino L, Jarvis A (2002) Computer tools for spatial analysis of plant genetic resources data: 2. FloraMap. Plant Genetic Resources Newsletter 130: 1-6.

Kiambi DK, Newbury HJ, Maxted N, Ford-Lloyd BV (2008) Molecular genetic variation in the African wild rice *Oryza longistaminata* A. Chev. et Roehr. and its association with environmental variables. African Journal of Biotechnology 7: 1446–1460.

Kim K-W, Chung H-K, Cho G-T, Ma K-H, Chandrabalan D, Gwag J-G, et al. (2007) PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. Bioinformatics 23: 2155–2162.

Kindt R, Coe R (2005) Tree Diversity Analysis. A Manual and Software for Common Statistical Methods for Ecological and Biodiversity Studies. World Agroforestry Centre (ICRAF), Nairobi.

Knight AT, Cowling RM, Rouget M, Balmford A, Lombard AT, et al. (2008) Knowing but not doing: selecting priority conservation areas and the research-implementation gap. Conservation Biology 22: 610-617.

Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. Meterologische Zeitschrift 15: 259-263.

Kozak KH, Graham CH, Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. Trends in Ecology and Evolution 23: 141–148.

Kumar P, Gupta VK, Misra AK, Modi DR, Pandey BK (2009) Potential of Molecular Markers in Plant Biotechnology. Plant Omics Journal 2: 141-162.

Labarta RA, Weber JC (1998) Valorización económica de bienes tangibles de cinco especies arbóreas agroforestales en la cuenca amazónica peruana. Revista Forestal Centroamericana 23: 12–21.

Leberg PL (2002) Estimating allelic richness: Effects of sample size and bottlenecks. Molecular Ecology 11: 2445-2449.

Leibing C, Signer J, van Zonneveld M, Jarvis A, Dvorak W (2013) Selection of provenances to adapt tropical pine forestry to climate change on the basis of climate analogs. Forests 4: 155-178.

Lek S, Guegan JF (1999) Artificial neural networks as a tool in ecological modelling, an introduction. Ecological Modelling 120: 65–73.

Lenihan JM, Neilson RP (1993) A rule-based vegetation formation model for Canada. Journal of Biogeography 20: 615–628.

Levis C, Souza PFd, Schietti J, Emilio T, Pinto JLPdV, et al. (2012) Historical human footprint on modern tree species composition in the Purus-Madeira interfluve, central Amazonia. PLoS ONE 7: e48559.

Liu C, Berry, PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28: 385 -393.

Lobell DB, Bänziger M, Magorokosho C, Vivek B (2011) Nonlinear heat effects on African maize as evidenced by historical yield trials. Nature Climate Change 1: 42–45.

Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17: 145–151.

Loiselle BA, Howell CA, Graham CH, Goerck JM, Brooks T, et al. (2003) Avoiding pitfalls of using species distribution models in conservation planning. Conservation Biology 17: 1591–1600.

Lora J, Herrero M, Hormaza JI (2009) The coexistence of bicellular and tricellular pollen in *Annona cherimola* Mill. (Annonaceae): Implications for pollen evolution. American Journal of Botany 96: 802-808.

Lora J, Herrero M, Hormaza JI (2011) Stigmatic receptivity in a dichogamous early-divergent angiosperm species, *Annona cherimola* Mill. (Annonaceae). Influence of temperature and humidity. American Journal of Botany 98: 265-274.

Lora J, Herrero M, Hormaza JI (2012) Pollen performance, cell number, and physiological state in the early-divergent angiosperm *Annona cherimola* Mill. (Annonaceae) are related to environmental conditions during the final stages of pollen development. Sexual Plant Reproduction 25: 157-167.

Lora J, Hormaza JI, Herrero M (2010) The progamic phase of an early-divergent angiosperm, *Annona cherimola* (Annonaceae). Annals of Botany 105: 221-231.

Lowe A, Harris S, Ashton P (2004) Ecological Genetics: Design, Analysis and Application. Blackwell Publishing, Oxford, UK.

Lowe AJ, Gillies ACM, Wilson J, Dawson IK (2000) Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis. Molecular Ecology 9: 831-841.

Lowe AJ, Boshier D, Ward M, Bacles CFE, Navarro C (2005) Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. Heredity 95: 255–273.

Magurran A (1988) Ecological Diversity and Its Measurement. Princeton University Press, Princeton, New Jersey.

Malhi Y, Aragao LEOC, Galbraith D (2009) Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. Proceedings of the National Academy of Sciences 106: 20610-20615.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology and Evolution 18: 189-197.

Mann CC (2005) In the land of the four quarters. In: 1491: new revelations of the Americas before Columbus. 1st edition, Knopf AA, US. 62-96.

Mannion AM (1999) Domestication and the origins of agriculture: an appraisal. Progress in Physical Geography 23: 37–56.

Bioversity International, INIA, LAFORGEN (2013) MAPFORGEN: Atlas para la conservación de los recursos genéticos forestales. <www.mapforgen.org>. Accessed at April 17th 2013.

Maxted N, Dulloo E, Ford-Lloyd BV, Iriondo JM, Jarvis A, et al. (2008) Gap analysis: a tool for complementary genetic conservation assessment. Diversity and Distributions 14: 1018–1030.

Maxted N, van Slageren MW, Rihan JR (1995) Ecogeographic surveys. In: Guarino L, Ramanatha Rao V, Reid R, editors. Collecting Plant Genetic Diversity. CABI International, Wallingford, UK. 255–285.

Maxted N, Kell SP (2009) Establishment of a Global Network for the In situ Conservation of Crop Wild Relatives: Status and Needs. Rome: FAO Commission on Genetic Resources for Food and Agriculture.

Maxted N, Iriondo JM, Ford-Lloyd BV, Dulloo ME, Kell SP, et al., editors (2008) Crop wild relatives conservation and use. CABI, London, UK.

Mayle FE (2004) Assessment of the Neotropical dry forest refugia hypothesis in the light of palaeoecological data and vegetation model simulations. Journal of Quaternary Science 19: 713–720.

McAleece N, Lambshead PJD, Paterson GLJ (1997) Biodiversity Professional. The Natural History Museum, London and the Scottish Association of Marine Science, Oban, Scotland.

McLeod MJ, Guttman SI, Eshbaugh WH (1982) Early Evolution of Chili Peppers (*Capsicum*) Economic Botany 36: 361-368.

McRae BH (2006) Isolation by resistance. Evolution 60: 1551–1561.

Meckelmann SW, Riegel DW, van Zonneveld M, Ríos L, Peña K, Ugas R, Quiñonez L, Mueller-Seitz E, Petz M (2013) Compositional characterization of native Peruvian chili peppers (*Capsicum* spp.). Journal of Agriculture and Food Chemistry DOI: 10.1021/jf304986q

Mercer KL, Perales HR (2010) Evolutionary response of landraces to climate change in centers of crop diversity. Evolutionary Applications 2010 3: 480-493.

Meyer RS, DuVal AE, Jensen HR (2012) Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. New Phytologist 196: 29–48.

Miller MP (2005) Alleles in space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. Journal of Heredity 96: 722-724.

Miller A, Schaal B (2005) Domestication of a Mesoamerican cultivated fruit tree, *Spondias purpurea*. Proceedings of the National Academy of Science 102: 12801-12806.

Miller AJ, Knouft JH (2006) Gis-based characterization of the geographic distributions of wild and cultivated populations of the mesoamerican fruit tree *Spondias purpurea* (Anacardiaceae). American Journal of Botany 93: 1757–1767.

Moat J (2007) Conservation assessment tools extension for ArcView 3.x, version 1.2. GIS Unit, Royal Botanic Gardens, Kew

Mora-Urpí J (1999) Origen y domesticación. In: Mora-Urpí J, Gainza EJ, editors. Palmito de pejibaye (*Bactris gasipaes* Kunth): su cultivo e industrialización. Editorial de la Universidad de Costa Rica, San José, Costa Rica. 17-24.

Mora-Urpí J, Weber JC, Clement CR (1997) Peach palm. Bactris gasipaes Kunth. Promoting the conservation and use of underutilized and neglected crops. No. 20. Institute of Plant Genetics and Crop Plant Research, Gatersleben/ International Plant Genetic Resources Institute, Rome, Italy.

Morcote-Rios G, Bernal R (2001) Remains of palms (Palmae) at archaeological sites in the New World: a review. Botanical Review 67: 309-350.

Moscone EA, Scaldaferro MA, Gabriele M,. Cecchini NM, Sánchez García Y, et al. (2007) The evolution of chili peppers (*Capsicum* – Solanaceae): a cytogenetic perspective. Acta Horticulturae 747: 137-170.

Motamayor JC, Lachenaud P, Da Silva e Mota JW, Loor R, Kuhn DN, et al. (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (Theobroma cacao L). PLoS ONE 3: e3311.

Myers N, Mittermeier RG, Mittermeier CG, da Fonseca GAB, Kent J, et al. (2000) Biodiversity hotspots for conservation priorities. Nature 403: 853-858.

National Research Council (1989a) Lost crops of the Incas: Little-known plants of the Andes with promise for worldwide cultivation, National Academy Press, Washington, DC.

National Research Council (1989b) Cherimoya. In: Lost crops of the Incas: Little-known plants of the Andes with promise for worldwide cultivation, National Academy Press, Washington, DC. 228-239.

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nature Review Genetics 12: 111-122.

Nelder J, Wedderburn R (1972) Generalized linear models. Journal of the Royal Statistical Society Series A 135: 370–384.

Newton AC, Allnut TR, Gillies ACM, Lowe AJ, Ennos RA, et al. (1999) Molecular phylogeography, intra-specific variation and the conservation of tree species. Trends in Ecology and Evolution 14: 140-145.

Newton AC, Oldfield S (2008) Red Listing the world's tree species: a review of recent progress. Endangered Species Research 6: 137-147.

Nic Lughadha E, Baillie J, Barthlott W, Brummitt NA, Cheek MR, et al. (2005) Measuring the fate of plant diversity: towards a foundation for future monitoring and opportunities for urgent action. Philosophical Transactions of the Royal Society B 360: 359–372.

Nichols KM, Neale DB (2010) Association genetics, population genetics, and conservation: revealing the genes underlying adaptation in natural populations of plants and animals. In: DeWoody JA, Bickman JW, Michler CH, Nichols KM, Rhodes OE, editors. Molecular approaches in natural resource conservation and management. Cambridge University Press. 123-168.

Ochoa C (1990) The Potatoes of South America: Bolivia. Cambridge University Press, Cambridge, UK.

Odong TL, van Heerwaarden J, Jansen J, van Hintum TJL, van Eeuwijk FA, et al. (2012) Statistical Techniques for Defining Reference Sets of Accessions and Microsatellite Markers. Crop Science 51: 2401-2411.

OECD (2003) Agriculture and Biodiversity: Developing Indicators for Policy Analysis. OECD Publishing.

Ortiz R, Sayre KD, Govaerts B, Gupta R, Subbarao GV, et al. (2008) Climate change: can wheat beat the heat? Agriculture, Ecosystems and Environment 126: 46–58.

Ouborg N, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. Trends in Genetics 26: 177-187.

Palmberge-Lerche C (2008) Thoughts on the conservation of forest biological diversity and forest tree and shrub genetic resources. Journal of Tropical Forest Science 20: 300-312.

Parra-Quijano M, Iriondo JM, de la Cruz M, Torres E (2011) Strategies for the development of core collections based on ecogeographical data. Crop Science 51: 656-666.

Pascual L, Perfectti F, Gutierrez M, Vargas AM (1993) Characterizing isozymes of Spanish cherimoya cultivars. HortScience 28: 845-847.

Patiño F, Cadima X (2009) Proyecto Parientes Silvestres - Conservación complementaria *in situ - ex situ* de papas silvestres. In Informes de Proyectos 2008 - 2009. Cochabamba, Bolivia: Fundación PROINPA.

Patiño F, Condori B, Segales L, Cadima X (2008) Distribución potencial actual y futura de especies silvestres endémicas de papa de bolivia. Revista de Agricultura 44: 37-44.

Patiño VM (2002) *Bactris gasipaes* Kunth. Historia y dispersión de los frutales nativos del neotrópico. Cali, Colombia. Centro Internacional de Agricultura Tropical, 2002. 464-537.

Pautasso M (2009) Geographical genetics and the conservation of forest trees. Perspectives in Plant Ecology, Evolution and Systematics 11: 157–189.

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288–295.

Pearce JL, Boyce MS (2006) Modelling distribution and abundance with presence-only data. Journal of Applied Ecology 43: 405–412.

Pearsall DM (1992) The origins of plant cultivation in South America. In: Cowan CW, Watson PJ, editors. The origins of agriculture: an international perspective. Smithonian Series in Archaeological Inquiry. 173-205.

Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12: 361–371.

Pekin BK, Boer MM, Macfarlane C, Grierson PF (2009) Impacts of increased fire frequency and aridity on eucalypt forest structure, biomass and composition in southwest Australia. Forest Ecology and Management 258: 2136-2142.

Penn MG, Moncrieff CB, Bridgewater SGM, Garwood NC, Bateman RM, et al. (2009) Using GIS techniques to model the distribution of the economically important xat palm *Chamaedorea ernesti-augusti* within the Greater Maya Mountains, Belize. Systematics and Biodiversity 7: 63-72.

Pennington RT, Prado DE, Pendry CA (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. Journal of Biogeography 27: 261–273.

Pennington RT, Lavin M, Prado DE, Pendry CA, Peel SK, et al. (2004) Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. Philosophical Transactions of the Royal Society of London B 359: 515–537.

Pennington RT, Lavin M, Särkinen T, Lewis GP, Klitgaard BB, et al. (2010) Contrasting plant diversification histories within the Andean biodiversity hotspot. Proceedings of the National Academy of Sciences 107: 13783–13787.

Pereira HM, Ferrier S, Walters M, Geller GN, Jongman RHG, et al. (2013) Essential biodiversity variables. Science 339: 227-278.

Perfectti F, Pascual L (1998) Characterization of cherimoya germplasm by isozyme markers. Fruit Varieties Journal 52: 53-62.

Perfectti F, Pascual L (2005) Genetic diversity in a worldwide collection of cherimoya cultivars. Genetic Resources and Crop Evolution 52: 959-966.

Perry L (2012) Ethnobotany. In: Russo VM, editor. Peppers, Production and Uses. CABI. 1-13.

Peters J (2008) Ecohydrology of wetlands: monitoring and modelling interactions between groundwater, soil and vegetation. PhD thesis, Ghent University, Ghent, Belgium.

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

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300: 1563-1565.

Petit RJ, Hu FS, Dicks CW (2008) Forests of the past: a window to future changes. Science 320: 1450-1452.

Phillips J (2009) A Brief Tutorial on Maxent. <www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc>. Accessed at April 17th 2013.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions Ecological Modelling 190: 231-259.

Pickersgill B (2007) Domestication of plant in the Americas: insights from Mendelian and molecular genetics. Annals of Botany 100: 925-940.

Pickersgill B, Heiser CB (1977) Origins, and distribution of plants domesticated in the New World tropics. In: Reed CA, editor. Origins of agriculture. Mouton, The Hague. 803-835.

Pinhasi R, Fort J, Ammerman AJ (2005) Tracing the origin and spread of agriculture in Europe. PLoS Biology 3: 2220-2228.

Popenoe W (1921) The native home of the cherimoya. Journal of Heredity 12: 331-336.

Pozorski T, Pozorski S (1997) Cherimoya and guanabana in the archaeological record of Peru. Journal of Ethnobiology 17: 235-248.

Prates-Clark CC, Saatchi SS, Agosti D (2008) Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data. Ecological Modelling 211: 309-323.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155: 945-959.

Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. Nature 457: 843-848.

R Development Core Team (2010) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org>. Accessed at November 26st 2012.

Ræbild A, Larsen AS, Jensen JS, Ouedraogo M, De Groote S, et al. (2011). Advances in domestication of indigenous fruit trees in the West African Sahel. New Forests 41: 297-315.

Rajora OP, Mosseler A (2001) Challenges and opportunities for conservation of forest genetic resources. Euphytica 118: 197–212.

Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, et al. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proceedings of the National Academy of Sciences 102: 15942-15947.

Ramanatha Rao V, Hodgkin T (2002) Genetic diversity and conservation and utilization of plant genetic resources. Plant Cell, Tissue and Organ Culture 68: 1–19.

Ramírez-Villegas J, Khoury C, Jarvis A, DeBouck DG, Guarino L, et al. (2010) A gap analysis methodology for collecting crop genepools: a case study with *Phaseolus* beans. PLoS ONE 5: e13497.

Rebelo AG, Sigfried WR (1992) Where should nature reserves be located in the Cape Floristic Region, South Africa? Models for the spatial configuration of a reserve network aimed at maximizing the protection of diversity. Conservation Biology 6: 243–252.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. Conservation Biology 17: 230-237.

Reis VM (2009) Relações genéticas entre raças e populações da coleção nuclear de pupunha (*Bactris gasipaes* Kunth) avaliadas com microssatélites. Master thesis, Universidade Federal do Amazonas.

Rodrigues DP, Filho SA, Clement CR (2004) Molecular marker-mediated validation of morphologically defined landraces of pejibaye (*Bactris gasipaes*) and their phylogenetic relationships. Genetic Resources and Crop Evolution 51: 871–882.

Romney AK, Weller SC, Batchelder WH (1986) Culture as consensus: a theory of culture and informant accuracy. American Anthropologist 88: 313-338.

Ronning CM, Kowalski SP, Sanford LL, Stommel JR (2000) Geographical variation of solanidane aglycone glycoalkaloids in the wild potato species *S. chacoense* Bitter. Genetic Resources and Crop Evolution 47: 359-369.

Rupprecht F, Oldeland J, Finckh M (2011) Modelling potential distribution of the threatened tree species *Juniperus oxycedrus*: how to evaluate the predictions of different modelling approaches? Journal of Vegetation Science 22: 647-659.

Russell J, Dawson IK, Flavell AJ, Steffenson B, Weltzien E, et al. (2011) Analysis of > 1000 single nucleotide polymorphisms in geographically matched samples of landrace and wild barley indicates secondary contact and chromosome-level differences in diversity around domestication genes. New Phytologist 191: 564-578.

Saénz-Romero C, Guzmán-Reyna RR, Rehfeldt GE (2006) Altitudinal genetic variation among *Pinus oocarpa* populations in Michoacán, Mexico: implications for seed zoning, conservation, tree breeding and global warming. Forest Ecology and Management 229: 340–350.

Salas A, Gaspar O, Rodríguez W, Vargas M, Centeno R, et al. (2008) Regeneration guidelines: wild potato. In: Dulloo ME, Thormann I, Jorge MA, Hanson J, editors. Crop specific regeneration guidelines [CD-ROM]. CGIAR System-wide Genetic Resource Programme, Rome, Italy.

Samuel AF (2012) Development of a Weitzman-type decision-support tool for the identification of cost-effective diversity-maximising crop genetic resources conservation programme strategies. Technical report, Bioversity International.

Sanchez PA, Ahamed S, Carré F, Hartemink AE, Hempel J, et al. (2009) Digital soil map of the world. Science 325: 680–681.

Santos RP, de Cristo-Araújo M, Picanço-Todrigues D, Filho SA, Clement CR (2011) Genetic variability and gene flow in hybrid and wild populations of peach palm accessed with RAPD markers. Revista Brasileira de Fruticultura 33: 1200–1208.

Scheldeman X (2002) Distribution and potential of cherimoya (Annona cherimola Mill.) and highland papayas (*Vasconcellea* spp.) in Ecuador. PhD Thesis, Ghent University.

Scheldeman X, Van Damme P, Ureña Alvarez JV, Romero Motoche JP (2003) Horticultural potential of Andean fruit crops exploring their centre of origin. Acta Horticulturae 598: 97-102.

Scheldeman X, Kanashiro M, Porro R, Dantas Medeiros R (2006) Taller sobre conservación y uso de frutales amazónicos en el ámbito de la Iniciativa Amazonica. Boa Vista, Roraima (Brasil), 27 a 29 de Septiembre 2006. Organized by IPGRI, EMBRAPA and the Amazon Initiative.

Scheldeman X, Willemen L, Coppens D, Eeckenbrugge G, Romeijn-Peeters E, Restrepo MT, et al. (2007). Distribution, diversity and environmental adaptation of highland papayas (*Vasconcellea* spp.) in tropical and subtropical America. Biodiversity and Conservation 16: 1867-1884.

Scheldeman X, van Zonneveld M (2010) Training manual on spatial analysis of plant diversity and distribution. Bioversity International, Rome, Italy.

Schroeder CA (1947) Hand pollination of cherimoya improves fruit set. California Avocado Society 32: 67-70.

Schueler S, Kapeller S, Konrad H, Geburek T, Mengl M, et al. (2012) Adaptive genetic diversity of forest trees: promise for future forests and a threatened resource - a case study on Norway spruce in Austria. Biodiversity and Conservation doi: 10.1007/s10531-012-0313-3.

Segarra-Moragues JG, Ojeda F (2010) Postfire response and genetic diversity in *Erica coccinea*: Connecting population dynamics and diversification in a biodiversity hotspot. Evolution 64: 3511-3524.

SERNAP (Servicio Nacional de Areas Protegidas Bolivia) (2011) Detalle de Areas Protegidas. Available: http://www.sernap.gob.bo.

Sheppard Jr GH, Ramirez H (2011) "Made in Brazil": human dispersal of the Brazil nut (*Bertholletia excelsa*, Lecythidaceae) in ancient Amazonia. Economic Botany 65: 44–65.

Silva CC (2004) Análise molecular e validação de raças primitivas de pupunha (Bactris gasipaes) por meio de marcadores RAPD. Masters Thesis, Universidade Federal de São Carlos/Universidade Federal do Amazonas, Manaus, Brazil.

Silvertown J (2009) A new dawn for citizen science. Trends in Ecology and Evolution 24: 467–471.

Simon R, Fuentes AF, Spooner DM (2011) Biogeographic implications of the striking discovery of a 4,000 kilometer disjunct population of the wild potato *S. morelliforme* in South America. Systematic Botany 36: 1062–1067.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82: 561–573.

Solano E, Feria TP (2007) Ecological niche modeling and geographic distribution of the genus *Polianthes* L. (Agavaceae) in Mexico: using niche modeling to improve assessments of risk status. Biodiversity and Conservation 16: 1885–1900.

Soldati M, Fornes L, van Zonneveld M, Thomas E, Zelener N (2013) An assessment of the genetic diversity of *Cedrela balansae* (Meliaceae) in Northwestern Argentina by means of combined use of SSR and AFLP molecular markers. Biochemical Systematics and Ecology 47: 45–55.

Sousa NR, Rodrigues DP, Clement CR, Nagao EO, Astolfi-Filho S, et al. (2001) Discriminação de raças primitivas de pupunha (*Bactris gasipaes*) na Amazônia brasileira por meio de marcadores moleculares (RAPDS). Acta Amazonica 31: 539-545.

Species Link (2012) < http://splink.cria.org.br>. Accessed at April 17th 2013.

Spooner D, Bamberg J (1994) Potato genetic resources: Sources of resistance and systematics. American Journal of Potato Research 71: 325-337.

Spooner DM, McLean K, Ramsay G, Waugh R, Bryan GJ, et al. (2005) A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. Proceedings of the National Academy of Sciences 102: 14694-14699.

Spooner DM, Salas A (2006) Structure, Biosystematics and Genetic Resources. In: Gopal J, Khurana SMP, editors. Handbook of potato production, improvement, and post-harvest management. Haworth's Press, Inc., Binghampton, New York. 1-39.

Spooner DM, van den Berg RG, Garcia W, Ugarte ML (1994) Bolivia potato germplasm collecting expeditions 1993, 1994: Taxonomy and new germplasm resources. Euphytica 79: 137-148.

Steiniger S, Bocher E (2008) An overview on current free and open source desktop GIS developments. International Journal of Geographical Information. Science 23: 1345–1370.

Steinmacher DA, Clement CR, Guerra MP (2007) Somatic embryogenesis from immature peach palm inflorescence explants: towards development of an efficient protocol. Plant Cell, Tissue and Organ Culture 89: 15-22.

Steinmacher DA, Guerra MP, Saare-Surminski K, Lieberei R (2011) A temporary immersion system improves in vitro regeneration of peach palm through secondary somatic embryogenesis. Annals of Botany 108: 1463–1475.

Stockwell D, Peters D (1999) The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographical Information Science 13: 143–158.

Svenning J-C, Skov F (2007) Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? Ecology Letters 10: 453-460.

Taberlet P, Zimmermann NE, Englisch T, Tribsch A, Holderegger R, et al. (2012) Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. Ecology Letters doi: 10.1111/ele.12004.

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 227: 1063-1066.

Tapia ME (2000) Mountain agrobiodiversity in Peru. Seed fairs, seed banks, and mountain-to-mountain exchange. Mountain Research and Development 20: 220–225.

Thomas E, van Zonneveld M, Loo J, Hodgkin T, Galluzzi G, et al. (2012) Present spatial diversity patterns of *Theobroma cacao* L. in the Neotropics reflect genetic differentiation in Pleistocene refugia followed by human-influenced dispersal. PLoS ONE 7: e47676.

Thomas M, Dawson JC, Goldringer I, Bonneuil C (2011) Seed exchanges, a key to analyze crop diversity dynamics in farmer-led on-farm conservation. Genetic Resources and Crop Evolution 58: 321–338.

Thormann I, Gaisberger H, Mattei F, Snook L, Arnaud E, et al. (2012) Digitization and online availability of original collecting mission data to improve data quality and enhance the conservation and use of plant genetic resources. Genetic Resources and Crop Evolution 59: 635–644.

Thuiller W (2003) BIOMOD: optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biology 9: 1353-1362.

Thuiller W, Lafourcade B, Engler B, Araújo MB (2009) BIOMOD: a platform for ensemble forecasting of species distributions. Ecography 32: 369-373.

Tognelli MF, Roig-Juñent SA, Marvaldi AE, Flores GE, Lobo JM, et al. (2009) An evaluation of methods for modelling distribution of Patagonian insects. Revista Chilena Historia Natural 82: 347-360.

Trognitz B, Scheldeman X, Hansel-Hohl K, Kuant A, Grebe H, et al. (2011) Genetic population structure of cacao plantings within a young production area in Nicaragua. PLoS ONE 6: e16056.

Tuberosa R, Graner A, Varshney RK (2011) Genomics of plant genetic resources: an introduction. Plant Genetic Resources 9: 151-154.

Tucker K, Rushton SP, Sanderson RA, Martin EB, Blaiklock J, et al. (1997) Modelling bird distributions: a combined GIS and Bayesian rule-based approach. Landscape Ecology 12: 77–93.

UNEP-WCMC (2010) Data Standards for the World Database on Protected Areas. UNEP-WCMC. < http://www.wdpa.org/>. Accessed at April 17th 2013.

Valenzuela CR, Padilla Suarez F (2002). A proposal for land use regulation of the Tunari national park, South Vertient. In: GISDECO 2002 proceedings: Governance and the use of GIS in developing countries, ITC, Enschede, 15-18 May 2002. 37-17.

Van Damme P, Scheldeman X (1999) Promoting cultivation of cherimoya in Latin America. Unasylva 198: 43-47.

Van Damme P, Van Damme V, Scheldeman X (2000) Ecology and cropping of cherimoya (Annona cherimola Mill.) in Latin America. New data from Ecuador. Fruits 55: 195-256.

Van Damme V, Gómez-Paniagua H, De Vicente MC (2011) The GCP molecular marker toolkit, an instrument for use in breeding food security crops. Molecular Breeding 28: 597-610.

van de Wouw M, Kik C, van Hintum T, Van Treuren R, Visser B, et al. (2010a) Genetic erosion in crops: concept, research results and challenges. Plant Genetic Resources 8: 1–15.

van de Wouw M, van Hintum T, Kik C, van Treuren R, Visser B, et al. (2010b) Genetic diversity trends in twentieth century crop cultivars: a meta analysis. Theoretical and Applied Genetics 120: 1241-1252.

van de Wouw M, van Treuren R, van Hintum T (2011) Authenticity of old cultivars in genebank collections: a case study on lettuce. Crop Science 51: 736–746.

van den Berg RG, Miller JT, Ugarte ML, Kardolus JP, Villand J, et al. (1998) Collapse of morphological species in the wild potato *Solanum brevicaule* complex (Solanaceae: Sect. *Petota*). American Journal of Botany 85: 92-109.

van Etten J (2006) Changes in farmers' knowledge of maize diversity in highland Guatemala, 1927/37-2004. Journal of Ethnobiology and Ethnomedicine 2:12 doi: 10.1186/1746-4269-2-12.

van Etten J (2011) Crowdsourcing crop improvement in sub-Saharan Africa: a proposal for a scalable and inclusive approach to food security. IDS bulletin 42: 102–110.

van Etten J, Hijmans RJ (2010) A geospatial modelling approach integrating archaeobotany and genetics to trace the origin and dispersal of domesticated plants. PLoS ONE 5: e12060.

van Heerwaarden J, Hellin J, Visser RF, van Eeuwijk FA (2009) Estimating maize genetic erosion in modernized smallholder agriculture. Theoretical and Applied Genetics 119: 875–888.

Van Leeuwen J, Lleras Pérez E, Clement CR (2005) Field genebanks may impede instead of promote crop development: Lessons of failed genebanks of “promising” Brazilian palms. Agrociencia 9: 61-66.

Van Soest L (2006) Global conservation strategy for potato. Global Crop Diversity Trust, Rome. <http://www.croptrust.org/documents/cropstrategies/Potato.pdf >. Accessed at April 17th 2013.

van Zonneveld M, Jarvis A, Dvorak W, Lema G, Leibing C (2009a) Climate change impact predictions on *Pinus patula* and *Pinus tecunumanii* populations in Mexico and Central America. Forest Ecology and Management 257: 1566–1576.

van Zonneveld M, Koskela J, Vinceti B, Jarvis A (2009b) Impact of climate change on the distribution of tropical pines in Southeast Asia. Unasylva 60: 24–28.

van Zonneveld M, Thomas E, Galluzzi G, Scheldeman X (2011) Chapter 15/16: Mapping the ecogeographic distribution of biodiversity and GIS tools for plant germplasm collectors. In: Guarino L, Ramanatha Rao V, Goldberg E, editors. Collecting Plant Genetic Diversity: Technical Guidelines - 2011 Update. Bioversity International, Rome, Italy.

van Zonneveld M, Scheldeman X, Escribano P, Viruel MA, Van Damme P, et al. (2012) Mapping genetic diversity of cherimoya (*Annona cherimola* Mill.): application of spatial analysis for conservation and use of plant genetic resources. PloS ONE 7: e29845.

Vanhove W, Van Damme P (2009) Marketing of cherimoya in the Andes for the benefit of the rural poor and as a tool for agrobiodiversity conservation. Acta Horticulturae 806: 497-504.

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants: selected writings of N.I. Vavilov. Ronald Press Company, New York.

Vavilov NI (1992a) The important agricultural crops of pre-Columbian America and their mutual relationship. In: Dorofeyev VF, editor. Origin and geography of cultivated plants. Cambridge University Press. 391-415.

Vavilov NI (1992b) Centers of origin of cultivated plants. In: Dorofeyev VF, editor. Origin and geography of cultivated plants. Cambridge University Press. 22-135.

Vavilov NI (1992c) The phyto-geographical basis for plant breeding. In: Dorofeyev VF, editor. Origin and geography of cultivated plants. Cambridge University Press. 316-366.

Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Sánchez JG, et al. (2008) Population structure and genetic diversity of new world maize races assessed by DNA microsatellites. American Journal of Botany 95: 1240–1253.

Viruel MA, Hormaza JI (2004) Development, characterization and variability analysis of microsatellites in lychee (*Litchi chinensis* Sonn., Sapindaceae). Theoretical and Applied Genetics 108: 896-902.

Vinceti B, Loo J, Gaisberger H, van Zonneveld MJ, Schueler S, et al. (2013) Conservation priorities for *Prunus africana* defined with the aid of spatial analysis of genetic data and climatic variables. PLoS ONE 8: e59987.

VMABCC (2009) Bioversity International. Libro Rojo de Parientes Silvestres de Cultivos de Bolivia. PLURAL Editores, La Paz, Bolivia.

Vranckx G, Jacquemyn H, Muys B, Honnay O (2011) Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation. Conservation Biology 26: 228-237.

Waltari E, Hijmans RJ, Peterson AT, Nyári SA, Perkins SL, et al. (2007) Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. PLoS ONE 2: e563.

Weber JC, Sotelo Montes C, Vidaurre H, Dawson IK, Simons AJ (2001) Participatory domestication of agroforestry trees: An example from the Peruvian Amazon. Development in Practice 11: 425 – 433.

Weitzman, M. L. (1998) The Noah's Ark Problem. Econometrica 66: 1279-1298.

Weller SC, Mann NC (1997) Assessing rater performance without a "gold standard" using consensus theory. Medical Decision Making 17: 71-79.

Widmer A, Lexer C (2001) Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. Trends in Ecology and Evolution 16: 267-269

Wiersum KF (1997) From natural forest to tree crops, co-domestication of forest and trees. Netherlands Journal of Agricultural Science 45: 425-438.

Willemen L, Scheldeman X, Soto Cabellos V, Salazar SR, Guarino L, et al. (2007) Spatial patterns of diversity and genetic erosion of traditional cassava (Manihot esculenta Crantz) in the Peruvian Amazon: an evaluation of socio-economic and environmental indicators. Genetic Resources and Crop Evolution 54: 1599–1612.

Willis KJ, Whittaker RJ (2002) Species Diversity - Scale Matters. Science 295: 1245-1248.

Willis F, Moat J, Paton A (2003) Defining a role for herbarium data in Red List assessments: a case study of *Plectranthus* from eastern and southern tropical Africa. Biodiversity and Conservation 12: 1537–1552.

Wolters B (1999) Zur Verbreitungsgeschichte und Ethnobotanik indianischer Kulturplanfzen, insbesondere des Kakaobaumes. Angewandte Botanik 73: 128-137.

Wolters B (2001) Dissemination of American economic plants on precolumbian sea routes by Amerindians. Migration and Diffusion l: 40-58.

Worthington M, Soleri D, Aragón-Cuevas F, Gepts P (2012) Genetic composition and spatial distribution of farmer-managed Phaseolus bean planting: an example from a village in Oaxaca, Mexico. Crop Science 52: 1721-1735.

Wright S (1943) Isolation by distance. Genetics 38: 114-138.

Zeballos H, Balderrama F, Condori B, Blajos J (2009) Economía de la papa en Bolivia (1998-2007). Cochabamba, Bolivia: Fundacion PROINPA.

Zeven AC (1999) The traditional inexplicable replacement of seed and seed ware of landraces and cultivars: a review. Euphytica 110: 181-191.

Zeven AC, De Wet JMJ (1982) Dictionary of cultivated plants and their regions of diversity: excluding most ornamentals, forest trees and lower plants. Centre for Agricultural Publishing and Documentation, Wageningen, The Netherlands.

Zhang D, Arevalo-Gardini E, Mischke S, Zúñiga-Cernades L, Barreto-Chavez A, et al. (2006) Genetic diversity and structure of managed and semi-natural populations of cocoa (*Theobroma cacao*) in the Huallaga and Ucayali valleys of Peru. Annals of Botany 98: 647–655.

Zohary D, Spiegel-Roy P (1975) Beginnings of fruit growing in the Old World. Science 187: 319-327.

Appendices

Appendix 3.1 The nine different variable combinations to develop species distribution models with Maxent.

Variable combination	19 bioclimatic variables*	4 bioclimatic variables**	Soil units	Ecological zones
1	X			
2		X		
3	X		X	
4	X			x
5			X	x
6		X	X	
7		X		x
8	X		X	x
9		X	X	x

*see www.worldclim.org for more details about the 19 bioclimatic variables

** annual mean temperature, annual precipitation, temperature seasonality, precipitation seasonality

Appendix 3.2 URL links to the Spanish questionnaire to validate species 'distribution models.

Annona cherimola: http://gisweb.ciat.cgiar.org/mapforgen/ann_che.html
Bactris gasipaes: http://gisweb.ciat.cgiar.org/mapforgen/bac_gas.html
Bertholletia excelsa: http://gisweb.ciat.cgiar.org/mapforgen/ber_exc.html
Cedrela odorata: http://gisweb.ciat.cgiar.org/mapforgen/ced_odo.html
Nothofagus alpina http://gisweb.ciat.cgiar.org/mapforgen/not_ner.html

Appendix 3.3 Expert scores and values of statistical-analysis based conventional parameters for cross-validation per model for each species.

Consensus-weighed expert scores

Variable combination	1	2	3	4	5	6	7	8	9	Median score	Range max- min score
A. cherimola	2.91	2.85	3.46	**3.90**	1.29	3.54	3.52	3.48	2.97	3.46	2.61
B. gasipaes	2.92	2.59	3.13	2.63	1.39	2.45	2.29	**3.18**	2.70	2.63	1.79
B. excelsa	3.54	**4.30**	3.02	2.28	2.19	3.93	3.05	3.00	2.51	3.02	1.69
C. odorata	3.11	**3.34**	2.95	2.48	1.78	2.17	1.95	2.61	1.91	2.48	1.55
N. alpine	2.54	2.79	2.77	2.83	1.30	3.80	**3.82**	3.12	3.35	2.83	2.52
Median score	2.92	2.85	3.02	2.63	1.39	**3.54**	3.05	3.12	2.70		
Range max-min score	1.00	1.71	0.69	1.62	0.90	1.76	1.87	0.87	1.44		

Un-weighed expert scores

Variable combination	1	2	3	4	5	6	7	8	9	Median score	Range max- min score
A. cherimola	2.67	**3.56**	3.22	3.00	1.78	3.00	2.56	3.00	2.67	3.00	1.78
B. gasipaes	2.44	2.22	2.89	2.78	2.11	2.33	2.44	**3.11**	2.56	2.44	1.00
B. excelsa	3.00	**3.20**	2.80	2.80	2.20	**3.20**	2.60	3.00	2.60	2.80	1.00
C. odorata	2.62	2.77	2.54	2.62	2.46	2.46	2.38	**2.85**	2.31	2.54	0.54
N. alpine	2.71	2.71	2.71	2.71	1.29	**3.57**	**3.57**	3.14	3.14	2.71	2.29
Median score	2.67	2.77	2.80	2.78	2.11	**3.00**	2.56	**3.00**	2.60		
Range max-min score	0.56	1.34	0.68	0.38	1.17	1.24	1.19	0.29	0.83		

Area Under Curve (AUC) of cross-validated models

Variable combination	1	2	3	4	5	6	7	8	9	Median score	Range max- min score
A. cherimola	0.963	**0.983**	0.967	0.978	0.891	0.976	0.975	0.976	0.965	0.975	0.085
B. gasipaes	0.844	0.779	0.857	**0.875**	0.601	0.738	0.758	0.87	0.786	0.758	0.269
B. excelsa	0.844	0.801	0.832	**0.889**	0.683	0.822	0.784	0.881	0.84	0.822	0.198
C. odorata	0.887	0.796	0.883	**0.901**	0.792	0.816	0.851	0.877	0.858	0.851	0.085
N. alpine	0.84	**0.889**	0.849	0.84	0.721	0.786	0.791	0.784	0.786	0.786	0.07
Median score	0.844	0.801	0.857	**0.889**	0.721	0.816	0.791	0.877	0.84		
Range max-min score	0.123	0.204	0.135	0.138	0.29	0.238	0.217	0.192	0.179		

Maximum values are in bold and underlined

Appendix 3.3 Continuation.

Maximum Kappa of cross-validated models

Variable combination	1	2	3	4	5	6	7	8	9	Median score	Range max-min score
A. cherimola	0.836	**0.917**	0.828	0.867	0.623	0.867	0.861	0.871	0.842	0.861	0.248
B. gasipaes	0.609	0.487	0.574	**0.6**	0.27	0.436	0.539	0.617	0.583	0.539	0.347
B. excelsa	0.653	0.547	0.627	**0.693**	0.427	0.599	0.667	0.653	0.68	0.653	0.253
C. odorata	0.593	0.463	0.607	**0.684**	0.489	0.509	0.596	0.642	0.605	0.596	0.153
N. alpine	0.600	**0.694**	0.635	0.663	0.400	0.682	0.529	0.565	0.682	0.565	0.282
Median score	0.609	0.547	0.627	**0.684**	0.427	0.599	0.596	0.642	0.68		
Range max-min score	0.243	0.454	0.254	0.267	0.353	0.431	0.332	0.306	0.259		

Commission error (%) of cross-validated models

Variable combination	1	2	3	4	5	6	7	8	9	Median score	Range max- min score
A. cherimola	7.63	3.97	7.42	4.25	25.37	5.12	5.93	**3.80**	6.25	5.93	21.57
B. gasipaes	39.39	34.78	34.78	**33.33**	40.25	41.48	45.45	35.90	42.68	41.48	9.56
B. excelsa	**39.13**	**39.13**	45.31	39.66	47.37	42.27	51.13	39.66	44.00	44.00	11.47
C. odorata	23.14	36.56	28.34	**21.63**	38.85	35.75	28.77	26.39	27.95	28.77	12.46
N. alpine	**16.67**	25.00	19.75	20.35	39.02	29.11	31.19	24.42	26.58	29.11	14.61
Median score	23.14	34.78	28.34	**21.63**	39.02	35.75	31.19	26.39	27.95		
Range max-min score	31.76	35.16	37.89	35.41	22.00	37.15	45.20	35.86	37.75		

Omission error (%) of cross-validated models

Variable combination	1	2	3	4	5	6	7	8	9	Median score	Range max- min score
A. cherimola	18.2	14.48	**12.74**	15.48	16.02	13.13	13.34	15.18	14.76	14.76	2.89
B. gasipaes	23.08	27.17	**14.49**	18.75	28.17	33.45	38.46	20.27	34.25	33.45	18.19
B. excelsa	**14.29**	**14.29**	22.73	14.71	29.41	18.25	58.82	14.71	20	20.00	44.11
C. odorata	20.50	21.35	**9.65**	13.89	19.08	20.48	16.20	12.98	16.87	16.87	7.49
N. alpine	25.51	27.78	22.47	**13.87**	21.28	30.34	16.39	23.81	20.99	21.28	13.94
Median score	20.50	21.35	**14.49**	14.71	21.28	20.48	16.39	15.18	20.00		
Range max-min score	11.22	13.49	13.08	4.88	13.39	20.32	45.48	10.83	19.49		

Maximum Kappa values with a bold and underlined font are the maximum values for a specific species and variable combination. In the case of commission and omission errors, minimum values are in bold and underlined font.

Appendix 3.4 Expert opinion per species (%) with respect to inclusion of areas where the species is absent (commission) in the model which they selected as best-fitting.

Weighed expert scores				Un-weighed expert scores		
Species	No	Yes	No answer	No	Yes	No answer
A. cherimola	3.98	40.91	55.11	11.11	33.33	55.56
B. excelsa	0.00	67.62	32.38	0	66.67	33.33
B. gasipaes	0.00	31.24	68.76	0	60	40
C. odorata	15.62	70.33	14.05	23.08	69.23	7.69
N. alpina	19.18	61.09	19.73	14.29	71.43	14.29
Mean	7.76	54.24	38.00	9.69	60.13	30.17

Expert opinion per species (%) about reasons for species absence in predicted areas of occurrence in the model which they selected as best-fitting.

Weighed expert scores				Un-weighed expert scores		
Species	Human disturbance	Outside distribution range	No answer	Human disturbance	Outside distribution range	No answer
A. cherimola	0.00	47.59	52.41	0	66.67	33.33
B. excelsa	14.92	61.28	23.79	33.33	66.67	16.67
B. gasipaes	0.00	0.00	100.00	0	66.67	33.33
C. odorata	68.57	31.43	0.00	55.56	44.44	0
N. alpina	24.04	75.96	0.00	20	80	0
Mean	21.51	43.25	35.24	20.11	63.22	16.67

Expert opinion per species (%) with respect to exclusion of areas where the species is present (omission) in the model which they selected as best-fitting.

Weighed expert scores				Un-weighed expert scores		
Species	No	Yes	No answer	No	Yes	No answer
A. cherimola	40.91	3.98	55.11	22.22	22.22	55.56
B. excelsa	36.28	53.40	10.33	44.44	44.44	11.11
B. gasipaes	0.00	27.42	72.58	0.00	60.00	40.00
C. odorata	56.68	21.22	22.10	53.85	38.46	7.69
N. alpina	30.75	49.52	19.73	42.86	42.86	14.29
Mean	32.92	31.11	35.97	32.67	41.60	25.73

Appendix 4.1 Differences between two taxonomies for Bolivian wild potato species

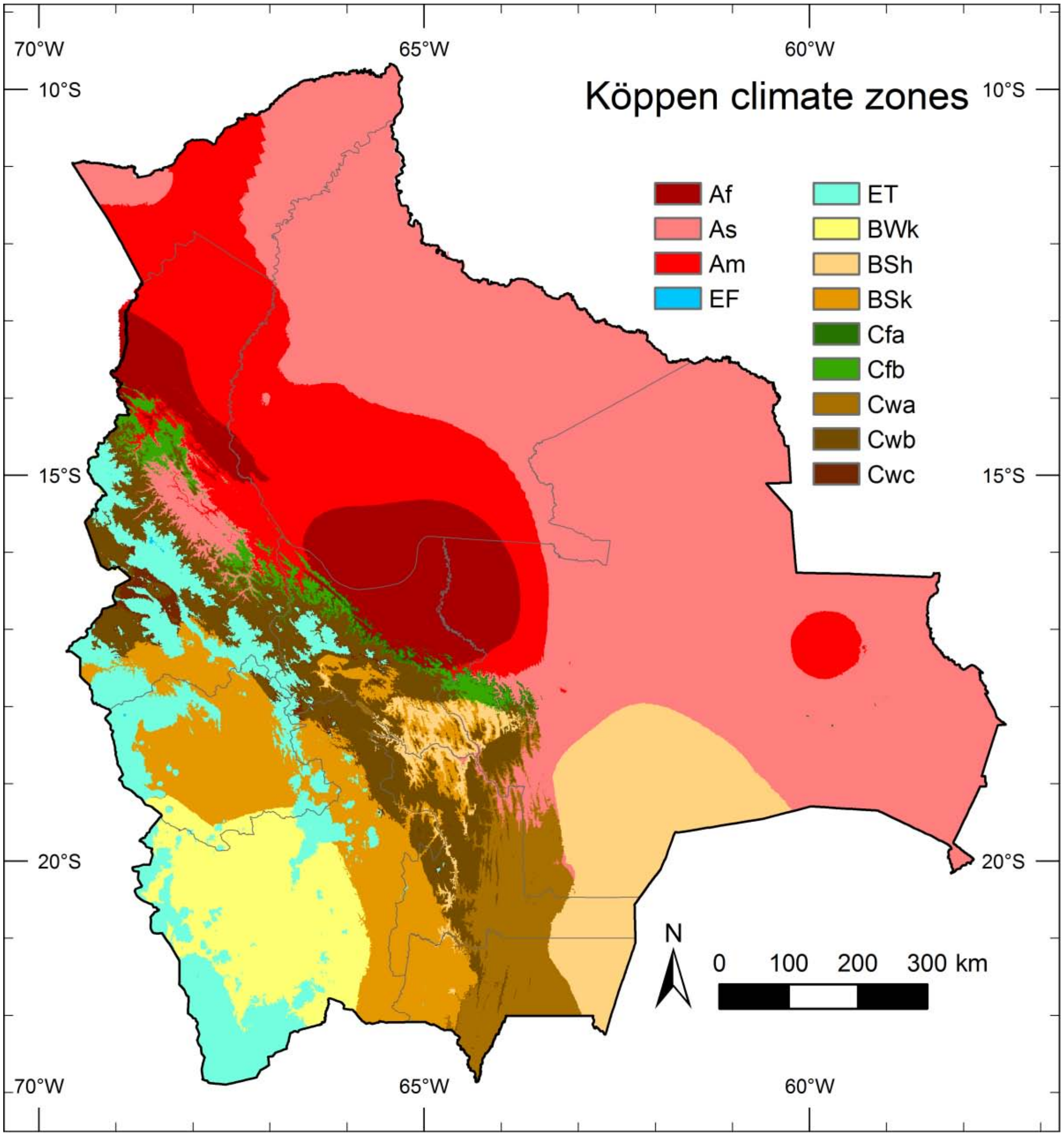
Nr.	Endemic wild potato taxa following Spooner and Salas (2006)	Nr.	Wild potato taxa following www.solanaceaesource.org	Endemic
1	*S. achacachense* Cárdenas		Synonym of *S. candolleanum* Berthault	
2	*S. alandiae* Cárdenas		Synonym of *S. brevicaule* Bitter	
3	*S. arnezii* Cárdenas		Synonym of *S. chacoense* Bitter	
4	*S. avilesii* Hawkes and Hjert.		Synonym of *S. brevicaule* Bitter	
5	*S. berthaultii* Hawkes	1	*S. berthaultii* Hawkes	*
6	*S. boliviense* Dunal	2	*S. boliviense* Dunal	*
7	*S. bombicynum* Ochoa	3	*S. bombicynum* Ochoa	*
8	*S. brevicaule* Bitter	4	*S. brevicaule* Bitter	*
9	*S. circaeifolium* Bitter	5	*S. circaeifolium* Bitter	*
10	*S. ×doddsii* Correl	6	*S. doddsii* Correl	*
11	*S. flavoviridens* Ochoa		Awaiting *S.* status designation	(*)
12	*S. gandarillasii* Cárdenas		Awaiting *S.* status designation	(*)
13	*S. hoopesii* Hawkes and K.A. Okada		Synonym of *S. brevicaule* Bitter	
14	*S. ×litusinum* Ochoa		Synonym of *S. berthaultii* Hawkes	
15	*S. neocardenasii* Hawkes and Hjert.	7	*S. neocardenasii* Hawkes and Hjert.	*
16	*S. neovavilovii* Ochoa	8	*S. neovavilovii* Ochoa	*
17	*S. soestii* Hawkes and Hjert.		Synonym of *S. circaeifolium* Bitter	
18	*S. ×sucrense* Hawkes		Synonym of *S. brevicaule* Bitter	
19	*S. ugentii* Hawkes and K.A. Okada		Synonym of *S. brevicaule* Bitter	
20	*S. violaceimarmoratum* Bitter	9	*S. violaceimarmoratum* Bitter	*
21	*S. virgultorum* (Bitter) Cárdenas and Hawkes		Synonym of *S. brevicaule* Bitter	

Appendix 4.2 Ploidy level and endosperm balance numbers (EBN) of the Bolivian endemic wild potato species.

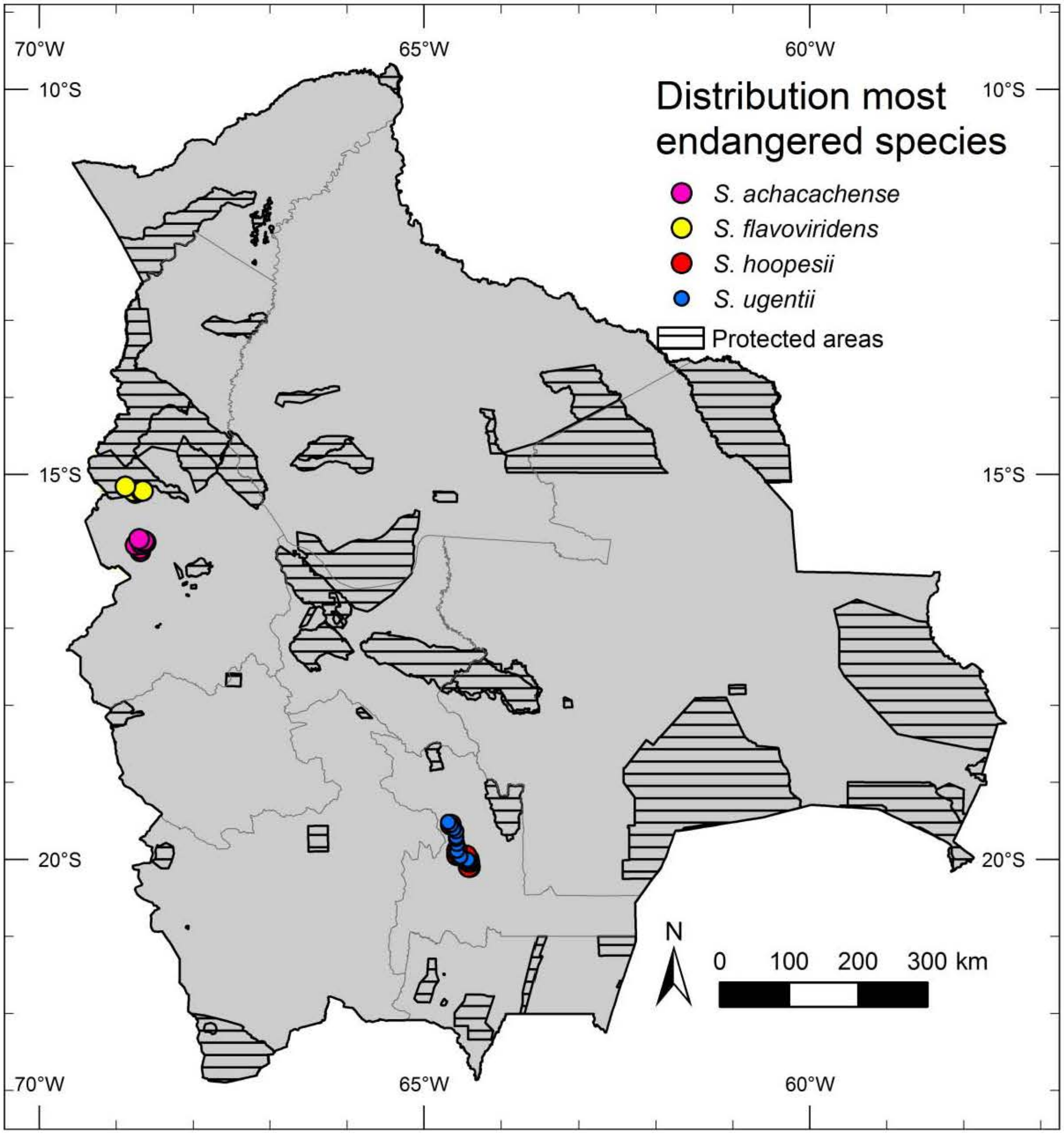
Species	Ploidy and (EBN)[*]
S. achacachense Cárdenas	2x
S. alandiae Cárdenas	2x
S. arnezii Cárdenas	
S. avilesii Hawkes and Hjrt.	2x
S. berthaultii Hawkes	2x (2EBN)
S. boliviense Dunal	2x (2EBN)
S. bombicynum Ochoa	4x
S. brevicaule Bitter	2x (2EBN)
S. circaeifolium Bitter	2x (1EBN)
S. x doddsii Correl (aln x chc)	2x (2EBN)
S. flavoviridens Ochoa	
S. gandarillasii Cárdenas	2x (2EBN)
S. hoopesii Hawkes and K.A. Okada	4x
S. x litusinum Ochoa (ber x tar)	2x (2EBN)
S. neocardenasii Hawkes and Hjert.	2x
S. neovavilovii Ochoa	2x (2EBN)
S. soestii Hawkes and Hjert.	2x
S. x sucrense Hawkes (adg x opl)	4x (4EBN)
S. ugentii Hawkes and K.A. Okada	4x
S. violaceimarmoratum Bitter	2x (2EBN)
S. vilgultorum (Bitter) Cárdenas and Hawkes	2x

[*] Ploidy and EBN determinations follow Spooner and Hijmans 2001. EBN refers to a genetic isolating mechanism that allows crosses between species with the same EBN and prevents crosses between different EBN groups (Hawkes 1990).

Appendix 4.3 Köppen climate classification on the basis of the criteria provided by Kottek et al. (2006) and calculated with 30-seconds resolution monthly precipitation and mean temperature data from Worldclim. Af = equatorial rainforest, fully humid; As = equatorial savannah with dry summer; Am = equatorial monsoon; EF = tundra climate; ET = frost climate; BWk = cold desert climate; BSh = hot steppe climate; BSk = cold steppe climate; Cfa = warm temperature climate, fully humid and hot summers; Cfb = warm temperature climate, fully humid and warm summer; Cwa = warm temperature climate, dry winter and hot summer; Cwb = warm temperature climate, dry winter and hot summer; Cwc = warm temperature climate, dry winter and cool summer.



Appendix 4.6 Distribution of most endangered wild potato species.



Appendix 4.5 Number of accessions per endemic wild species conserved *ex situ* in genebanks according to the potato germplasm conservation strategy (van Soest 2006) and updated with new accessions collected by PROINPA.

Species	INTA	BOL	CIP	PI	CGN	CPC	IPK	VIR	POL	CZE	Sum
S. achacachense		4		1	4		1				10
S. alandiae		20	15	17	13	2	8	6			81
S. arnezii		7		6	2		4				19
S. avilesii		17	3	3	3		3	5			34
S. berthaultii	1	31	33	62	34	12	12	41	1	1	228
S. boliviense	13	23	10	25	25	6	14	25			141
S. bombycinum											0
S. brevicaule	1	15	9	27	14	2	5	15			88
S. circaeifolium		20	9	15	16	3	11	7			81
S. flavoviridens		4									4
S. gandarillasii		11	1	7	3	3	5	6			36
S. hoopesii		9	2	8	4		2				25
S. neocardenasii		4	1	2	1	1	2	2			13
S. neovavilovii		2									2
S. soestii		1					1				2
S. ugentii		3	2	5	3		2				15
S. violaceimarmoratum		8	8	8	5	1	4	7			41
S. virgultorum		6	1		7	1	2	1			18
S. ×doddsii		2	2	13	3	2	4	5			31
S. ×litusinum											0
S. ×sucrense		48	20	40	52	10	8	15			193
Total:	15	235	116	239	189	43	88	135	1	1	1062

Where INTA= Estación Experimental Balcarce -Instituto Nacional de Tecnología Agropecuaria, Argentina; BOL=Bolivian potato collection; CIP=International Potato Center, Peru; PI= Potato Introduction Project, USA; CGN=Centre for Genetic Resources, Netherlands; CPC=Common Wealth Potato Collection, UK; IPK=Institute of Plant Genetic Resources and Crop Plant Research, Germany; VIR= Vavilov Research Institute of Plant Industry, Russia; POL=Plant Breeding and Acclimatization Institute, Poland; CZE=Potato Research Institute, Czech Republic.

Curriculum vitae

Professional interest

- Diversity analysis to enhance use of plant genetic resources for sustainable agricultural development.
- Crop and tree species diversification as risk management tool for climate change adaptation.
- Monitor status and trends of dynamics in economic plant diversity and use.
- Geospatial diversity and distribution analysis to support biodiversity conservation strategies.

Personal information

Full name: Maarten Jan van Zonneveld
Date of birth: 11 December 1979
Nationality: Dutch
Civil state: married, two daughters
Residence: Palmira, Colombia
Post address:
Bioversity International,
Recta Cali-Palmira km 17,
Cali, Colombia
Tel: +57 2 4450048 ext. 118
Mobile: +57 317 5595083
Work email: mvanzonneveld@cgiar.org
Personal email: mvzonneveld@gmail.com

Academic formation

2011 to 2013	*PhD candidate in Applied Biological Sciences.* Ghent University, Belgium. Title: Economic plant diversity and distribution: application of spatial analysis for the conservation of germplasm diversity in Latin America.
1998 to 2005	*Degree equivalent to MSc in Forest and Nature Conservation.* Wageningen University & Research (UR), the Netherlands. *MSc thesis:* Field study about the effects of positive plant interactions on plant diversity along an arid scrubland-temperate rainforest boundary, National Park Fray Jorge, IV region, Chile. University of La Serena, Chile and Wageningen UR.

MSc internship: Field assistant in the improvement of production systems for small-scale farmers. Agraría (NGO), VIII region, Chile.

Work experience

2009 to date	*Associate scientist in diversity analysis for conservation and use.* Bioversity International, Regional Office of the Americas, Cali, Colombia. Bioversity is the former International Plant Genetic Resources Institute (IPGRI).
2006 to 2009	*Associate expert in conservation and use of forest genetic resources.* Bioversity International, regional office of the Americas, Cali, Colombia.

Research activities

2012 to date	Co-developer of a strategic action plan to enhance conservation and use of plant genetic resources for the adaptation of Mesoamerican agricultural systems to climate change.
2010 to date	Coordinator of characterization and evaluation activities to select promising chili pepper varieties for high-value product development in Bolivia and Peru.
2010 to date	Selection of barley and durum wheat varieties for local adaptation of agricultural systems in Ethiopia to global climate change.
2008 to date	Geospatial analysis of taxonomic and genetic diversity for *in situ* conservation of several economic tree species and crop wild relatives.
2007 to date	Coordinator of MAPFORGEN: a website to share knowledge about tree species diversity and distribution. www.mapforgen.org
2006 to 2009	Climate change impact and adaptation studies on the distribution of subtropical pine tree species and productivity of subtropical pine forests.

Communication and outreach

2006 to date	Participation in regional and global meetings and congresses to enhance the agenda of plant genetic resources within the context of sustainable use and conservation.
2006 to 2010	Coordinator of LAFORGEN, the Latin American Forest Genetic Resources Network.

Capacity building

2007 to date	Development of training material on spatial analysis of plant diversity and distribution and more than 20 training courses given in Latin America, Europe and Africa.

Language proficiency

- Dutch (mother tongue), Spanish (excellent), English (excellent)

Software/computer skills

- Statistics and modelling: SPSS, R, InfoStat, Sigmaplot, GenAlEx, InfoStat, Maxent, Structure.
- GIS: ArcGis, ArcView, DIVA-GIS.
- Non-Specific Software: Windows, MS office, experienced in Excel, familiar with Access.

Additional courses

- Sustainable management of tropical forests, field practical at the experimental station Caparo. Universidad de los Andes, Mérida, Venezuela. October 2008.
- Forest genetics. Postgraduate course. Universidad de Buenos Aires, Bariloche, Argentina. October 2008.
- Biotechnological data analysis. Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) Turrialba, Costa Rica. August 2007.

Reviewer (since 2001)

- Acta Physiologiae Plantarum, Applied Vegetation Science, Conservation Genetics, Journal of Biogeography, Journal of Ecology, Journal of Vegetation Science, PLOS ONE, Revista Fitotecnia Mexicana.

Peer-reviewed articles

Vinceti B, Loo J, Gaisberger H, **van Zonneveld M**, Schueler S, Konrad H, Kadu CAC, Geburek T (2013) Conservation priorities for *Prunus africana* defined with the aid of spatial analysis of genetic data and climatic variables. PLoS ONE 8: e59987. http://dx.plos.org/10.1371/journal.pone.0059987

Leibing C, Signer J, **van Zonneveld M**, Jarvis A, Dvorak W (2013) Selection of provenances to adapt tropical pine forestry to climate change on the basis of climate analogs. Forests 4: 155-178. http://www.mdpi.com/1999-4907/4/1/155

Meckelmann SW, Riegel DW, **van Zonneveld M**, Ríos L, Peña K, Ugas R, Quiñonez L, Mueller-Seitz E, Petz M (2013) Compositional characterization of native Peruvian chili peppers (*Capsicum* spp.). Journal of Agriculture and Food Chemistry DOI: 10.1021/jf304986q. http://pubs.acs.org/doi/abs/10.1021/jf304986q

Soldati M, Fornes L, **van Zonneveld M**, Thomas E, Zelener N (2013). An assessment of the genetic diversity of *Cedrela balansae* (Meliaceae) in Northwestern Argentina by means of combined use of SSR and AFLP molecular markers. Biochemical Systematics and Ecology 47: 45–55. http://www.sciencedirect.com/science/article/pii/S0305197812002360

Graefe S, Dufour D, **van Zonneveld M**, Rodriguez F, Gonzalez A (2013). Peach palm (*Bactris gasipaes*) in tropical Latin America: implications for biodiversity conservation, natural resource management and human nutrition. *Biodiversity and Conservation* 22 (2): 269-300. http://rd.springer.com/article/10.1007/s10531-012-0402-3

Thomas E, **van Zonneveld M**, Loo J, Hodgkin T, Galluzzi G, van Etten J (2012). Present-day spatial diversity patterns of *Theobroma cacao* L. in Tropical America reflect a combination of long-standing human influences superimposed on genetic differentiation in Pleistocene refuges. PLoS ONE 7: e47676. http://dx.plos.org/10.1371/journal.pone.0047676

van Zonneveld M, Scheldeman X, Escribano P, Viruel MA, Van Damme P, Garcia W, Tapia C, Romero J, Sigueñas M, Hormaza JI (2012). Mapping genetic diversity of cherimoya (*Annona cherimola* Mill.): application of spatial analysis for conservation and use of plant genetic resources. PLoS ONE 7: e29845. http://dx.plos.org/10.1371/journal.pone.0029845

van Zonneveld M, Gutierrez JR, Holmgren M (2012). Shrub facilitation increases plant diversity along an arid scrubland-temperate rainforest boundary in South America. Journal of Vegetation Science 23 (3): 541–551. http://onlinelibrary.wiley.com/doi/10.1111/j.1654-1103.2011.01379.x/abstract

van Zonneveld, M, Jarvis A, Dvorak W, Lema G, Leibing C (2009). Climate change impact predictions on *Pinus patula* and *Pinus tecunumanii* populations in Mexico and Central America. Forest Ecology and Management 257: 1566-1576.
http://dx.doi.org/10.1016/j.foreco.2008.12.027

Leibing C, **van Zonneveld, M**, Jarvis A, Dvorak W (2009). Adaptation of tropical and subtropical pine plantation forestry to climate change: realignment of *Pinus patula* and *Pinus tecunumanii* genotypes to 2020 planting site climates. Scandinavian Journal of Forest Research 24: 483-493.
http://www.tandfonline.com/doi/abs/10.1080/02827580903378642

van Zonneveld M, Koskela J, Vinceti B, Jarvis A (2009). Impact of climate change on the distribution of tropical pines in Southeast Asia. Unasylva 231/232: 24-29.
ftp://ftp.fao.org/docrep/fao/011/i0670e/i0670e05.pdf

Meckelmann S, Riegel D, Avila T, Bejarano C, **van Zonneveld M**, Petz M (2013) Untersuchung von 114 nativen bolivianischen Chili-Sorten auf bioaktive und wertgebende Inhaltstoffe [Investigation of 114 native Bolivian chili accessions for bioactive and valuable components] *Lebensmittelchemie, accepted*

Peer-reviewed book chapters and manuals

van Zonneveld, M, Thomas E, Galluzzi G, Scheldeman X (2011) Chapter 15/16: Mapping the ecogeographic distribution of biodiversity and GIS tools for plant germplasm collectors. In: Guarino, L., Ramanatha, Rao. V., Goldberg, E., editors. 2011. *Collecting Plant Genetic Diversity: Technical Guidelines - 2011 Update*. Bioversity International, Rome, Italy.
http://cropgenebank.sgrp.cgiar.org/index.php?option=com_content&view=article&id=390&Itemid=557&lang=english

Scheldeman X, **van Zonneveld M (**2010). Training manual on spatial analysis of plant diversity and distribution. Bioversity International. Rome, Italy. pp. 179.
http://www.bioversityinternational.org/training/training_materials/gis_manual.html

Spanish version of training manual available at:
http://www.bioversityinternational.org/nc/publications/publication/issue/manual_de_capacitacion_en_analisis_espacial_de_diversidad_y_distribucion_de_plantas.html

French version of training manual available at:
http://www.bioversityinternational.org/index.php?id=19&user_bioversitypublications_pi1[showUid]=7010

Presentations at congresses and symposia

Van Zonneveld M. Estrategia Mundial para la Conservación y Uso de Recursos Fitogenéticos. 1ª Feria de la Agrodiversidad y Agroproductos. Parque Ecológico Xochitla, Tepotzolán, México. 7-8 November 2012. Invited speaker.

van Zonneveld M, Ramirez M, Williams D, Petz M, Meckelmann S, Avila T, Bejarano C, Rios L, Jaeger M, Libreros D, Amaya K, Scheldeman X. Underutilized Capsicum pepper diversity in its Andean centre of origin: advances in high-value differentiation. *6th International Crop Science Congress.* Bento Gonçalves, Brazil, 6-10 August 2012.

van Zonneveld M, Castañeda N, Van Damme V, Thomas E, Guevara E, Scheldeman X. MAPFORGEN: Compartir conocimiento para conservar nuestra diversidad genética forestal. *VII Simposio Internacional de Recursos Genéticos de América Latina y el Caribe SIRGEALC*. 21-23 Noviembre de 2011. Resúmenes de los trabajos presentados. Quito Ecuador. ISBN 978-9942-07-215-3.

van Zonneveld MJ, Jarvis A, Dvorak W, Koskela J, Vinceti B, Snook L. Impact of climate change on distribution and performance of tropical pine species [in Central America and Southeast Asia]. *Adaptation of Forests and Forest Management to Changing Climate with Emphasis on Forest Health: A Review of Science, Policies and Practices*. Umeå, Sweden, 25-28 August 2008. Book of Abstracts and Preliminary Programme. pp. 257.

van Zonneveld M. Conservación de poblaciones naturales de pinos centroamericanos en el contexto del cambio climático. Bosques Tropicales y Desarrollo, Seminario Internacional. Jardín Botánico de Medellín Joaquin Antonio Uribe, Medellín, Colombia. 18-21 November 2008. Invited speaker.

van Zonneveld M. El impacto del cambio climático en las fuentes de semilla: Un estudio de caso sobre las procedencias de *Pinus patula* y *Pinus tecunumanii* en América Central. *Memorias del Simposio de Recursos Genéticos para América Latina y el Caribe.* Cd. de México noviembre 2007. Universidad Autónoma de Chapingo. pp. 217-219.

Additional conference papers

Leibing C, Signer J, **van Zonneveld M**, Jarvis A, Dvorak W. Selection of provenances to adapt tropical pine forestry to climate change on the basis of climate analogues. Breeding for Value in a Changing World, Jacksonville, Florida USA, February 4 – 7, 2013

Meckelmann S, Riegel D, Avila T, Bejarano C, **van Zonneveld M**, Petz M. Bioactive and valuable compounds in 114 native Bolivian chili accessions. 2012. International pepper proceedings 21st Int. Pepper Conference 2012 Naples/Florida, 4-6 November. pp. 38
http://www.conference.ifas.ufl.edu/pepper2012/Proceedings.pdf

Bejarano Martínez C, Equise H, Mayan C, Matthias J, **van Zonneveld M**, Amaya K. Bio-inputs: A Clean Alternative for Bolivian Chili Export. 2012. International pepper proceedings 21st Int. Pepper Conference 2012 Naples/Florida, 4-6 November. pp 24 http://www.conference.ifas.ufl.edu/pepper2012/Proceedings.pdf

Avila T, Atahuachi M, Reyes X, Claure T, van Zonneveld M. Collection, taxonomic identification and cropping methodologies development for some species of *Capsicum* in Bolivia. International pepper proceedings 21st Int. Pepper Conference 2012 Naples/Florida, 4-6 November. pp 52
http://www.conference.ifas.ufl.edu/pepper2012/Proceedings.pdf

Libreros D, Ríos Lobo L, **van Zonneveld M**, Williams D. Conservando la integridad genética: el uso de jaulas de exclusión en la multiplicación de germoplasma de *Capsicum* en el Perú. *58th Annual Meeting of the Interamerican Society for Tropical Horticulture*. Lima, Perú, 3-6 Septiembre de 2012.

Jalonen R, **van Zonneveld M**, Thomas E, Gaisberger H, Vinceti B, Hong LT, Loo J. Identifying tree populations for conservation action through geospatial analyses. In Multinational and Transboundary Conservation of Valuable and Endangered Forest Tree Species.Guangzhou (China), 5-7 Dec 2011. IUFRO p. 98-101 ISBN:978-3-902762-15-3

Thomas E, **van Zonneveld M**, Signer J. MAPFORGEN: El caso de la diversidad intra-específica de *Theobroma cacao* L. En América Latina. *VII Simposio Internacional de Recursos Genéticos de America Latina y el Caribe SIRGEALC.* 21-23 Noviembre de 2011. Resúmenes de los trabajos presentados. Quito Ecuador. ISBN 978-9942-07-215-3.

Jäger M, **van Zonneveld M**, Scheldeman X, Ramírez M. Vincular bancos de germoplasma y pequeños productores con mercados de alto valor – el ejemplo de la diversidad de *Capsicum* en Perú y Bolivia. *VII Simposio Internacional de Recursos Genéticos de América Latina y el Caribe SIRGEALC.* 21-23 Noviembre de 2011. Resúmenes de los trabajos presentados. Quito Ecuador. ISBN 978-9942-07-215-3.

Williams DE, Ríos Lobo L, **van Zonneveld M.** La colección nacional de *Capsicum* respalda el aprovechamiento de ajíes en el Perú. *VII Simposio Internacional de Recursos Genéticos de America Latina y el Caribe SIRGEALC.* 21-23 Noviembre de 2011. Resúmenes de los trabajos presentados. Quito Ecuador. ISBN 978-9942-07-215-3.

Jäger M, Scheldeman X, **van Zonneveld M.** Linking Gene Banks and Small Farmers to High Value Markets - The Example of *Capsicum* Diversity in Peru and Bolivia. *Proceedings of the Tropentag.* 2010 Congress, Zurich, Switzerland.

Marchelli P, Azpilicueta MM, **van Zonneveld M**, Gallo LA. Evolutionary potential of two ecologically contrasting southern beeches to face future climatic change. in Vinceti B., Neate P. (comps.). *Conference on 'Forest ecosystem genomics and adaptation*. San Lorenzo de El Escorial (Madrid), Spain, 9-11 June 2010. Book of abstracts. Bioversity International (Rome, Italy) and INIA (Madrid, Spain). pp. 206.

Soldati MC, Fornes L, **van Zonneveld M**, Zelener N. 2010. Diversidad genética molecular de C*edrela balansae* (Meliaceae) en las yungas (selva Tucumano-Boliviana) del noroeste Argentino. B12P -0004. *Congreso Latinoamericano (IV Argentino) de Conservación de la Biodiversidad*. 2010 - San Miguel de Tucumán – Argentina.

van Zonneveld M, Scheldeman X, Ramirez M, Quintana E. Impacto del trabajo en redes: ejemplos de casos exitosos en las Américas. Instituto de Investigaciones Agropecuarias (INIA), Ministerio de Agricultura, Chile. *Proceeding tomo 1, VII Simposio de Recursos Genéticos para América Latina y el Caribe*. Pucón Chile, 28 al 30 de octubre 2009. pp. 121-123.

Scheldeman X, **van Zonneveld M**, Castañeda N, Jarvis A. Recursos fitogenéticos y cambio climático: oportunidad y retos. Instituto de Investigaciones Agropecuarias (INIA), Ministerio de Agricultura, Chile. *Proceeding tomo 1, VII Simposio de Recursos Genéticos para América Latina y el Caribe*. Pucón Chile, 28 al 30 de octubre 2009. pp. 39-42.

Leibing C, **van Zonneveld M**, Jarvis A, Dvorak B. Adaptation of Tropical and Subtropical Pine Plantation Forestry to Climate Change: Climate Proofing Seed Material of *Pinus patula* and *Pinus tecunumanii.* Tropentag, October 6-8, 2009, Hamburg, Germany.

Leibing C, **van Zonneveld M**, Jarvis A. Adaptation of tropical pine plantation forestry to climate change. *Conference Series: Earth and Environmental Science 6. Climate Change: Global Risks, Challenges and Decisions*. Session 38 Adapting Forests to Climate Change, 10–12 March 2009, Copenhagen, Denmark.

van Zonneveld M, Salcedo J, Baena M. LAFORGEN una red para promover los recursos genéticos forestales en América Latina. *Congreso Internacional de Desarrollo, Medio Ambiente, y Recursos naturales. Volumen III*. Cochabamba, Bolivia, 11-13 julio 2007. Universidad Mayor de San Simón. pp. 1647.