ASSESSING PLANT COLLECTION EFFORT IN THE GUÁNICA FOREST RESERVE

By

Maricelys Figueroa Ruíz

A thesis submitted in partial fulfillment

of the requirement for the degree of

MASTER IN SCIENCE

In

BIOLOGY

UNIVERSITY OF PUERTO RICO

MAYAGUEZ CAMPUS

2015

Approved by:

______________________________	______________________________
Duane A. Kolterman, PhD Member, Graduate Committee	Date
______________________________	______________________________
Jeanine Vélez - Gavilán, MSc. Member, Graduate Committee	Date
______________________________	______________________________
Jesús D. Chinea, PhD President, Graduate Committee	Date
______________________________	______________________________
Jorge L. Almodóvar - Montañez , PhD Representative of Graduate Studies	Date
______________________________	______________________________
Nanette Diffoot - Carlo, PhD Chairperson of the Department	Date

Abstract

Museum data provide information on the distribution of species documented as presences. These data are vital in the development of herbaria through the world promoting scientific research in STEM (Science, Technology, Engineering and Mathematics, by its acronym in English) disciplines and requires high reliability. However, the data collection effort could be biased by taxon, site, time or environment, which could prevent researchers from answering questions correctly, or what is worse, give false answers. In this project we assessed the representativeness of the plant collection effort in the Guánica Forest Reserve (GFR), Puerto Rico. We measured the significance of the species, spatial, temporal and environmental bias in a database for this area through the use of Geographic Information Systems (GIS) and statistical analysis. Analyzing a total of 1,807 records georeferenced in ArcMap 10, we identified the presence of each bias in the collection, with the taxonomical and temporal bias resulting as the most significant. We suggest some strategies for collecting during the first months of the year in less visited areas in order to improve the quality of the collection. We hope that this study can be executed in other regions of importance for the Island's species and resources conservation. This analysis can serve as a guide for future explorations and as the basis for the use of this plant collection in studies based on GIS.

Resumen

Los datos de museo proveen información sobre la distribución de las especies en la forma de una presencia documentada. Éstos datos se han convertido en vital importancia en el desarrollo de los herbarios, impulsando la investigación científicas dentro de las disciplinas “STEM” (Ciencia, Tecnología, Ingeniería y Matemáticas, por sus siglas en inglés) y requieren un alto grado de confiabilidad. Sin embargo, el esfuerzo de colección podría estar sesgado taxonómica, espacial, temporal o ambientalmente, lo que podría prevenir a los investigadores de contestar sus preguntas correctamente o lo que es aún peor, dar contestaciones falsas. En este estudio, evaluamos la representatividad del esfuerzo de colección de plantas en la Reserva del Bosque de Guánica (RBG), Puerto Rico. Se midió la significancia del sesgo espacial, taxonómico y temporal en una base de datos para la zona utilizando Sistemas de Información Geográfica (SIG) y análisis estadístico. Analizando un total de 1,807 registros en ArcMap10, identificamos la presencia de cada uno de los sesgos en la colección, siendo los más marcados, el sesgo taxonómico y el temporal. Sugerimos aumentar el esfuerzo de colecta durante los primeros meses del año en las áreas menos visitadas para mejorar la calidad de la colección. Esperamos que este estudio pueda ejecutarse en otras regiones de importancia en la conservación de especies y recursos naturales en la Isla. Este análisis podría servir de guía para futuras exploraciones y como base para el uso de esta colección de plantas en estudios basados en SIG.

.

Dedicatoria

A mis hijas y a mis dos sobrinas por ser mi motivo principal para superarme y querer ser mejor. A mi esposo por su apoyo y aguante todo este tiempo. A mi mamá y a mi papá por todo su apoyo, por demostrarme en todo momento lo que es el amor incondicional. A mis hermanas y a cada una de esas mujeres importantes de mi familia por simplemente estar ahí. Y por último, a Dios; que esta tesis es para tu gloria Dios por permitirme vivir esta etapa de superación y por cada una de las oportunidades que pusiste en mi camino.

Acknowledgments

I want to say thanks to all the people that helped me during my preparation:

To all the members of my graduate committee for their guidance and patience in this journey: Dr. Chinea for the opportunities provided and teaching me to work with new technologies, Dr. Kolterman for all the history and knowledge you passed to me, and Jeanine Vélez for all your help with BRAHMS.

To NASA-EPSCoR project's "Hyperspectral imaging for biodiversity assessment in a coastal and terrestrial ecosystem" for funding a big part of my studies. To Dr. Mei Yu, from the Institute for Tropical Ecosystem Studies" and the United States Fish and Wildlife Service for the internship opportunities. And special thanks to Dr. Dimuth Siritunga, as a member of the Howard Hughes Medical Institute, and the University of Puerto Rico at Mayagüez for sponsored my professional trips to the Botanical Society of America meeting in New Orleans and to the Caribbean Biodiversity Congress in Dominican Republic.

To all of the professors from the University of Puerto Rico at Mayagüez who crossed my path for some reason, thanks for leaving a mark.

To Omar Monsegur for the free consultancy, the time and the patience. For the opportunities and all the knowledge given to me; I am very grateful.

To special people I met during these years: María Méndez, Mary Jiménez and Vilmarie Rivera for being so special women, without their help the slope would have been even steeper. Donato Seguí for giving me so many lessons, not necessarily related to ecology. Thanks for being such a wonderful human being. May God keep blessing you all.

To the family of graduate students I found in the Biology building. There are so many great friends, and even though I did not mention all, I do want to say special thanks to my angels Jessica López, Laura Vázquez and Roxanne Almodóvar because without you I would never have passed the Biometrics with A and B. Thanks for the coffee, the study sessions and the encouraging words. Love you with all my heart.

Finally, thanks to my whole family, my center and my north. For calming my tears when the pressure attacked and reminding me how good I am.

Thanks to all for enriching my brain with knowledge and my life with survival skills, and for being part of the process of becoming who I am today.

Table of contents

List of tables

List of figures

List of appendices

Introduction

In recent years, the use of museum data in scientific research has increased significantly. First of all, museum data provide information about the species that is being used in projects about Genetics, Taxonomy, Biogeography, Ecology and other fields in Biology (Graham *et al*. 2004, Hijmans *et al*. 2000, Newbold 2010). Second, in some cases, this information is being captured in electronic databases, and could be available through the Internet. Therefore, it represents very important data easily available to the scientific community (Graham *et. al* 2004, Newbold 2010, Rivers *et. al* 2011, Schdmit-Lebuhn *et. al* 2013). Graham *et al*. (2004) discussed about the unique combination of attributes that museum data provide: (1) massive information with associated collecting events describing time and place; (2) records of species backed by preserved organisms that enable the individual verification of the species' identity; (3) records or field notes that add considerable value to the specimens themselves; (4) historical distribution of the organisms which provides a platform for the assessment of biodiversity dynamics; (5) taxonomically current data increasing research on phylogenetic, environmental analysis and comparative genomics.

However, a museum collection consists of specimens collected by different people based on different objectives and if the data are not properly assessed their use in scientific research could be limited. Because the data are dependent on the sampling method, the information provided by the collector can be deviated to one variable more than another; resulting in a taxonomic, temporal, spatial or environmental bias (Hortal *et al*. 2007). Researchers are responsible on assessing the data for these biases to prevent wrong interpretation of it (Rich and Woodruff 1992, Schdmit-Lebuhn *et. al* 2013).

The Guánica Forest Reserve (GFR), located in the south of Puerto Rico, has been intensively surveyed after 1919, when it was declared a forest reserve (Álvarez-Ruíz *et al.* 1990, DNER 1981). Due to its protection, it is now one of the best remnants of subtropical dry forest in the Caribbean and its vegetation has recovered remarkably, becoming the habitat of 460 plant species, including endemics and endangered species. Most of the GFR plant collection is maintained in the Herbarium of the Biology Department at the University of Puerto Rico in Mayagüez (MAPR) and is electronically available through the Botanical Research and Herbarium Management System (BRAHMS).

Because of the importance of the GFR herbarium data in the conservation of the natural resources and species on the Island, the objective of this project was to assess the representativeness of the forest plant collection effort and give suggestions to improve its quality.

Literature Review

Museums, private natural history collections, and herbaria store preserved organisms that are used in scientific research. Besides providing us information about the distribution of species in the form of recorded occurrences, herbarium data provide massive information associated with time and place, preserved organism that enable the species identity verification, field notes, historical distribution and taxonomically current data increasing research on phylogenetic, environmental analysis and comparative genomics (Graham *et al*. 2004). Herbarium data provide the basis for different ecological projects (Graham *et al*. 2004, Hijmans *et al*. 2000, Newbold 2010). For example, Sousa-Baena *et al.* (2013) assessed digital data available in the *species*Link network to represent gaps in current knowledge about Brazilian floristic composition. They concluded that existing knowledge can guide biodiversity surveys and inventory efforts. However, due to its importance in the environmental, ecological and conservation issues, special attention is needed to enhance collections data. The Finnish Museum of Natural History in Finland is working with the development of biological atlases for the European vascular flora based primarily on herbarium collections. Until 2013 they had published 16 volumes of 20% of the flora and almost 5,000 digital maps (Finnish Museum of Natural History 2014). In Burkina Faso, Africa, Schmidt *et al.* (2005) worked with herbarium data to model plant diversity maps of four different families and concluded that digitized species occurrence data, linked to environmental data, and modeling algorithms should be an important advantage for the documentation, assessment and conservation of the biodiversity. They also developed maps to guide future collection efforts according to the species and the environment. In 2011, Rivers *et al*. used herbarium records to determine a preliminary conservation status of 661 endemic species of Leguminosae and Orchidaceae from Madagascar and concluded that despite the fact that the majority of world's plants have poorly known distributions represented by few specimens, herbarium records can still be used to make robust preliminary assessments. Species occurrence data are increasingly being captured in electronic databases and are being made available worldwide, receiving the name of biodiversity databases. Two examples are the Global Biodiversity Information Facility (GBIF

2001) and the Biodiversity Monitoring and Assessment Project (BioMAP–Egypt 2005), used by Tim Newbold, in 2010, to assess the applications and limitations of the museum data for Conservation and Ecology. Databases are created for different purposes. For example, BIOTA-Canarias (1999) was designed to manage specimen-based, spatially and taxonomically referenced data for researchers' use (Hortal *et al*. 2007). The BRAHMS database was designed to manage and integrate data and images of specimens, botanical studies, field observations, live collections, seed banks and literature (Schmidt *et al.* 2005).These kind of database has become a primary tool in ecological research, increasing the use of collection data in a wide range of projects, as previously mentioned. These biodiversity databases store information contained on the specimen labels, such as taxonomical identification, growth status and, sometimes, community composition. They also include information on the locality where the specimen was collected, which further expands their potential applications (Pearson 2008; Soberón *et al*. 2000). However, herbarium data depend on the collector's preferences or the project's objectives, and may be skewed. If they are, then the scientific value of a project is questionable, or at least the collection could be unrepresentative of the natural distribution of the organisms (Hijmans *et al.* 2000).

Species are not uniformly distributed. Biotic and abiotic factors will lead a species' populations to distribute themselves in an aggregated or random pattern in time and space. Because the distribution of organisms is not known prior to data analysis, successful collecting may depend on finding populations from the full range of environmental and geographical variability to which the species is adapted (Brown and Marshall 1995, Hijmans *et al.* 2000). But such variability may not be represented in a collection. The collection can deviate from the ideal sample for a number of reasons, for example, limits on accessibility, time, and funding. In addition, the collection may be biased if the specimens were collected with a specific objective other than biodiversity assessment, as happens in most cases. According to the Merriam Webster Dictionary (2015) the term bias may refer to "a systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others; a deviation of the expected value of a statistical estimate from the quantity it estimates". Botanists usually collect to rescue endangered species, to identify a specific trait or for fulfilling a specific research need. Therefore, sampling effort is limited, scattered and not standardized, and the inventories could be biased toward easily accessible sampling sites (Graham *et al*. 2004, Hijmans *et al*. 2000, Newbold 2010).

Distribution data for many of the species in biodiversity databases are prone to different kinds of bias (Hijmans *et al.* 2000, Graham *et al*. 2004, Stockwell and Peterson 2002). Schmidt-Lebuhn *et al.* (2013) used the Asteraceae specimen data from the Australia's Virtual Herbarium to test for species bias. They found different levels of bias according with the phenological trait assessed and concluded that extra care should be taken when relying on specimen databases for studies examining species distributions, patterns of biodiversity, or the rarity of species. Soberón and Peterson (2004) addressed the applications of plants' biodiversity informatics and the limitations bias could present when answering biogeographical questions, concluding that detection of the limitations depend on the participation of an active community of taxonomists to improve herbarium data uses and understanding. Kadmon *et al.* (2004) used the floristic database of the Israel Nature and Park Authorities to investigate the effect of roadside bias on the accuracy of predictive maps for woody plants in Israel produced by bioclimatic models. Their results supports the impact of bias in the model's predictions and they discussed theoretical and practical considerations to correct the bias in biodiversity databases.

Bias in biodiversity databases may compromise the description of biodiversity patterns from the raw information compiled in them. The most common bias present in the work of botanists is the sampling bias. A sampling bias is a consistent systematic error that arises from the sample selection, thus reducing the accuracy of the object sampled. The sampling bias can be further classified into spatial, taxonomic, temporal, and environmental bias (Newbold 2010; Schdmit-Lebuhn *et al*. 2013; Soberón *et al*. 2000). Hijmans *et al*. (2000) evaluated the representativeness of a wild potato genebank collection database in Bolivia and defined and assessed the presence of four types of sampling bias: species, species-area and infrastructure bias. Two of these biases are spatial: species-area and infrastructure bias. The species bias, is a taxonomical bias. Researchers found the presence of all of these biases in the Bolivian wild potato collection compromising the existing diversity or the actual geographic representativeness of the Intergenebank Potato Database.

The **taxonomic bias** refers to unequal collection among taxa in an area. It might result from differences in the probability of finding one species relative to other species. It could also result because of differences in abundance or because of the collector's preferences for a particular species due to a specific trait or research needs (Hijmans *et al*. 2000). In some cases it could be justified, for example, if the collector is focused on endangered or invasive species. However, it could compromise an accurate description of biodiversity patterns (Hortal *et al.* 2007).

In Australia, Schdmit-Lebuhn *et. al* (2013) assessed the taxonomical bias for Asteraceae's specimen data to test the over and under-representation of plants with specific characteristics and as expected a bias was detected in different levels. They concluded that extra care should be taken when working with specimen databases.

Hijmans *et al*. (2000) assessed the taxon bias in Bolivian wild potato collection. Treating all species, subspecies and varieties in the database as separate taxa, the authors tabulated the number of accessions per taxa (A_s). To assess the taxonomic bias with a Chi – square test, the observed distribution of A_s was compared to the expected distribution if all accessions were equally allocated among all species. The collection was dominated by a few taxa indicating a strong taxon bias.

The **spatial bias** is an asymmetry of the representation of spatial information. Newbold (2010) states that historical samplings have clearly been spatially biased, particularly in tropical and arid environments. According to Hijmans *et al*. (2000) the **species–area bias** refers to over-sampling or under-sampling of a species in relation to the size of the area in which it occurs. Theoretically, to maximize collection diversity, the number of records collected per species should increase with the size of the area in which it is found, assuming a larger area implies a proportionate increase in species diversity. Therefore, a species-area bias exists if a collection contains too many accessions for some species, and hence too few for others, relative to the size of the areas in which the species appear. Hijmans *et al*. (2000) evaluated this bias for a wild potato gene bank collection in Bolivia through different statistical methods. In 100 × 100 km grid cells, they determined the number of grid cells in which each species of wild potato was collected and with a linear and logarithmic regression looked for a relationship between the number of accessions per species *(A_s)* and the number of grid cells *(G_s)* in which each species *(S)* was present . Then, a Chi – square test was performed to compare the observed relationship between A_s and G_s versus the expected relationship, if all accessions were equally allocated to each species based on the number of grid cells in which it occurs. Despite the high correlation value obtained, Hijmans *et al*. (2000) identified the presence of the spatial bias in the collection, because the relationship did not increase proportionally among species.

Infrastructure bias is also spatial. It refers to oversampling near roads and towns. For reasons of efficiency, logistics and convenience, collectors tend to collect close to roads, trails and accessible sites. The effort of a random sampling must be uniform over different distances from the main roads identified. If it is not, an infrastructure bias is present. Even though in practice explorers cannot sample randomly, due to time limitation or accessibility, infrastructure bias can affect the accurate representation of the species distribution (Hijmans *et al*. 2000, Kadmon *et al*. 2004).

To assess the infrastructure bias for the wild potato collection, Hijmans *et al.* (2000) divided Bolivia in 1 × 1 km grid cells and calculated the distance from the genebank data's collection points to the nearest roads and to the nearest towns. Then, with a Chi – square test the authors compared the relationship between the observed distribution and the expected distribution, if all accessions were equally collected at random distances. The obtained distribution contrasted sharply with the random distribution plotted by the authors, indicating a significant infrastructure bias.

Kadmon *et al*. (2004) assessed the roadside bias with an index corresponding to the probability to get "*n"* points (out of *N*) near the road, when the probability to get a single point is known. In addition the distribution of the collection sites was compared to a null model of random distribution using a Kolmogorov-Smirnov test, obtaining a P-value < 0.0001 that strongly confirms the existence of an infrastructure bias.

The **temporal bias** results when the samples are not collected uniformly over time. It refers to oversampling within certain periods. Natural factors, such as rain or wind or personal factors, like the collector's preference or schedule, influence the time when most of records are collected. In the case of plant specimens, one factor likely to influence collectors' activities is the known or suspected phenological condition of the plants. If the collection has too many or too few accessions in a certain period of time, the collection is biased.

Peterson *et al*. (1998) analyzed the temporal bias for a bird collection in Mexico available in a large database assembled by the authors. They divided the study area in grid cells of 1° (100 × 100 km approximately) and plotted the records by century and by seasons of the year. The 19th Century presents few and sparsely distributed records, however numerous additional bird specimens reside in European museums that were only partially recorded in their database. The 20th Century showed great improvements in the precision of localities and quality of ancillary information of specimens, presenting a more complete coverage; but a good coverage only for certain parts of Mexico. When comparing the collection effort through the seasons of the year, the bias was more evident, presenting a different distribution among seasons with the best coverage during the spring, followed by summer, winter and autumn. The seasonal distribution was also related to the regions of the area. Seasonal pattern was expected due to the periodic influxes of migrants as well as seasonal changes in behavior, ecology, habitat use and freshness of plumage.

Soberón *et al*. (2000) worked with butterfly collection, also in Mexico. They grouped the records by decades and analyzed the resulting distribution as a function of the absolute number of records and the accumulated fraction of records. As a result, the collection effort was not regularly distributed over time;

showing a peak of collecting effort for different taxa in periods of time when experts in a given taxonomic group were most active. Therefore the authors concluded that the collection is temporally biased. Because botanists collect to fulfill a specific research needs (Hijmans *et al*. 2000), we expect a similar pattern in plant collecting; with a peak in collecting effort during months when reproductive material is available or when the collectors are more active.

The **environmental bias** is an asymmetry of the representation of the environmental information. It refers to sampling some habitats more than others. Data on the environment, species and species assemblages are needed to characterize areas and prioritize them for conservation (Williams *et al*. 2002). Features such as vegetation type or environmental classes are likely to be the best surrogates for analyses at a large portion of overall biodiversity (Margules and Pressey 2000).

Many applications of herbarium data involve analyses of the environment that species inhabit, for example, to predict species distribution (Loiselle *et al*. 2008) or to map plant diversity (Schdmit *et al.* 2005). For these applications, gaps in the spatial coverage of records may not be a problem as long as the data are not environmentally biased. Kadmon *et al.* (2004) found that the accuracy of distribution models for woody plants in Israel was decreased by the spatial bias and not by the environmental gradient, because the distribution of roads in Israel was not environmentally biased. This means that roads in Israel were constructed through the entire gradient of the climatic variable used in the study. It is assumed that a biodiversity database should have specimens collected in all possible environmental gradients, but yet these assumptions about the primary data are often untested and violating these assumptions will compromise model predictions (Loiselle *et al*. 2008).

In some cases spatial bias may result in environmental bias. For example, Hortal *et al*. (2007) assessed the environmental bias based on the spatial bias. To demonstrate the presence of the spatial bias in the biodiversity database BIOTA for a seed-plant collection in Tenerife, authors used a Spearman's correlation analysis between the number of records and the number of species observed per grid cell to identify those areas where the inventories could have been saturated and used it as data to generate a predictive model for species richness. They obtained predictive maps highly unreliable with high values of explained variability of the data, but at the same time, high values of prediction error and low values of predictive power, preventing accurate representation of the spatial patterns of richness variation. The authors developed a protocol to assess data quality in order to improve model predictions and obtain reliable information on the geographic distribution of biodiversity.

These limitations can hinder the usefulness of a collection in a database, even if all the data available were gathered exhaustively (Hortal *et al.* 2007). Researchers need to know the limitations a database can present and how to overcome them to achieve better results.

Puerto Rico is located in a biodiversity hotspot and its flora has been surveyed since 19th Century, even though the scientific education was not promoted under the Spanish regimen. Although most of the collecting efforts in Puerto Rico were done after the Spanish American War, in 1898, the biological assessment began earlier; for example, Paul Sintenis collected in the mid 1800's for Ignatious Urban Flora Portoricensis. From 1913 to 1970, the New York Academy of Sciences, in association with the Department of Agriculture of Puerto Rico and the University of Puerto Rico initiated the publication of the journal "Scientific Survey of Porto Rico and the Virgin Islands", focused on themes related to botany, geology and archeology. This event allowed the entrance of different collectors to survey the vegetation of the new territories of the United States, being Dr. N. L. Britton, from the New York Botanical garden, the most notable collector in the Island during the first three decades of the 20th Century. In the 1960s, Dr. Henry Alain Liogier arrived to the Island to work with his study about the Flora of Puerto Rico and the Virgin Islands, which was published in 5 volumes from 1985 to 1997. However he moved to Dominican Republic during the 1970's where he founded the "Herbarium of Dominican Republic" and formally inaugurated the "National Botanical Garden". It is not until the 1980's that the arrival of new botanists (Dr. Pedro Acevedo, Dr. Gary Breckon, and Dr, Franklin Axelrod in the 1990's) boosts the plant collection activity in the Island (Kolterman, personal communication).

The Island possesses two of the Biosphere Reserves declared by the United Nations Educational Scientific and Cultural Organization (UNESCO 2011): El Yunque National Forest, a subtropical rain forest designated in 1976; and the GFR, a subtropical dry forest designated in 1981. Tropical dry forests are the most threatened of the tropical forest types (Janzen 1988). Much of the dry forests in the Caribbean have been cut, primarily for agriculture, urbanization and fuel. Furthermore, dry forests tend to be favored for human habitation, and as a result only a few tropical and subtropical dry forests remain undisturbed (Murphy and Lugo 1990). The GFR was subjected to an intense human activity, like housing, logging and agriculture. According to Molina–Colón and Lugo (2006), 23 families lived in El Maniel sector before it became part of the forest in 1948, maintaining small farmlands; pasture goats, horses, cattle, and other domestic animals; and harvesting green tree stems and branches for charcoal production. The forest terrains were also used for forest plantations and more recently for passive recreation and conservation of its flora and fauna (Chinea 1990). From 1922 to 1970, the United States Forest Service (USFS) conducted an improvement tree planting program to counteract the human impact on the reserve.

The GFR is a natural area comprising the best remnant of dry forest vegetation in the Caribbean and has been the focus of worldwide research on the ecology of tropical dry forest since the 1970's. By 1990, 191 projects were documented for the forest, approximately 51% of them about the flora (Álvarez-Ruíz *et al.* 1990). It is now the largest Commonwealth protected area over limestone substrate (Lugo *et al.* 1996) with an estimated area of about 4,480 ha (11,400 acres). The conservation activities began since 1919, when the GFR was declared a state forest (DNER 1981). From 1950 to 1980 the GFR was maintained as a conservation unit due to geographical and economic reasons because the hotel, industrial and urban expansions were mainly in the northern area of the island (Canals 1990). The environmental conservation at a government level began during the creation of the Commonwealth Environmental Quality Board and the Department of Natural and Environmental Resources (DNER). The recognition received in 1981 as a biosphere reserve increased the vision of the GFR as an ecological jewel of the same magnitude as El Yunque National Forest (Lugo 1990).

According to Monsegur (2009), the GFR flora comprises approximately 460 species (Appendix I). Monsegur also documented rare, endangered and endemic species: 20 species were documented as rare, for example, *Metastelma monense* and *Zephyranthes proctorii*; 47 species as critical elements by the DRNA and seven species (*Catesbaea melanocarpa, Cordia rupicola, Eugenia woodburyana, Mitracarpus maxwelliae, Mitracarpus polycladus, Ottoschulzia rhodoxylon* and *Trichilia triacantha)* as threatened/endangered by the United States Fish and Wildlife Service (USFWS). Of the 460 confirmed species, 19 were documented as restricted to the Puerto Rican archipelago and another 16 species restricted to the Puerto Rican bank, of which three are restricted to the forest. Additionally, Monsegur informed about the probability of extirpation of some species from the forest, because since 1960 or before, they have not been collected or seen in the reserve. Part of the GFR collection is maintained in the Herbarium of the Biology Department, at the University of Puerto Rico in Mayaguez (MAPR) and available in its database. This database includes data from other herbaria in Puerto Rico and United States.

Despite the long protection status of the GFR, the loss of the species that seems to be extirpated may signal the impacts on biodiversity due to former land use of the forest and raised questions about the forest's plant collection. To what extent is this collection representative of the biodiversity in the forest? Are all species represented? Do those areas which present a higher number of collection sites have more species richness than the areas that were under-collected? Are those species that have not been documented in the last six decades really extirpated or were the habitats where they would normally be found not searched?

When herbarium records are used to assess biodiversity, it is important to understand the level of representativeness of the real situation in the study area. To assess if this plant collection is representative of the biodiversity of the GFR, different kinds of biases were evaluated: species bias, species-area, infrastructure, temporal and environmental bias.

Objectives

This study assessed the plant collection efforts in the Guánica Forest Reserve. The more specific objectives of this study were:

1. To georeference plant specimen localities documented for the forest.
2. To describe the collecting biases for the GFR plant collection.
3. To propose collection strategies aimed at improving the quality of the collection.

Methods

Study Area - The GFR is located on the southwest coast of Puerto Rico within portions of the Guánica, Guayanilla, and Yauco municipalities separated by the Guánica Bay into two units, East and West Units (Figure 1). Several small islets also belong to this reserve but they were not included in this study. Its extent is about 4,480 ha and it is located roughly at latitude 17°58´ N and longitude 66°55´ W (DRNA, 1981). According to Ewel and Whitmore (1973) this area has been classified as "Dry Forest on limestone" and as "Subtropical Dry Forest", which is characterized by low annual rainfall and a high evapotranspiration ratio. Murphy and Lugo (1990) discussed the composition of the forest. It receives approximately 800 mm of precipitation per year concentrated mostly during the months of July to November, and with a marked dry season during January to March. Moisture availability as a function of shallow soils, plus low rainfall and its seasonality are the factors suggested as determining forest productivity, growth characteristics, water loss and physiognomy.

The floristic structure and composition of the overall forest is affected by different factors, such as disturbance history, elevation, aspect, substrate and water availability. The plant diversity in the GFR can also be explained by the wide diversity of habitats produced by the proximity of the limestone basement, to the surface and the variations in soil depth (Murphy and Lugo 1990). In addition, the variable topography with a mixture of hills, deep canyons or ravines, gentle to steep slopes and the effect of airborne salt contribute to the diversity of habitats (Monsegur 2009). The land use is another important factor that affects the flora biodiversity in the GFR.

Several decades of recovery after the cessation of land uses had taken place for over a century in the GFR. The species composition of the recovered mature forests is different from those present in pre-conversion forests and after 100 years of development, the original species had not return in the emerging forest; being altered by the presence of some persistent exotic species, such as *Leucaena leucocephalla, Swietenia mahogany* and *Megathyrsus maximus* (Chinea 1990; Molina-Colón and Lugo 2006; Wadsworth

1990). This sustains that the change in species composition seems to be the most significant effect of human use and landscape modification in the GFR (Molina-Colón and Lugo 2006).

BRAHMS – The Botanical Research and Herbarium Management System was created at the University of Oxford, UK, in 1985, to manage and integrate data and images from specimens, botanical surveys, field observations, living collections, seed banks and literature, optimizing its use for the widest possible range of curation and research services and outputs. BRAHMS management system is being used by projects in over 60 countries worldwide, some based in larger botanical institutes and gardens and other in small herbaria, university departments and field stations, being the National Herbarium of the Netherlands the largest single database with some 3 million specimens as of 2014 (University of Oxford 2015).

The Herbarium of the Biology Department of the University of Puerto Rico at Mayaguez (MAPR) is using BRAHMS since 1998, adding more than 46,000 records from different parts of Puerto Rico, for example, the floras of Mona Passage, Caja de Muertos, Vieques and GFR, among others. Another reason to use this software is its flexibility to integrate efficient data and images promoting research in plant systematics and plant diversity interpretation. In addition, the successful introduction and use of BRAHMS database at MAPR provides information of specimens collected in Puerto Rico by other herbaria, researchers and projects, converting MAPR into a fairly complete source of information for some groups of plants in Puerto Rico (University of Oxford 2015, J. Vélez personal communication).

Georeferencing Records. As mentioned before, most of the plants collected at the GFR have been deposited in the MAPR, but additional data from other herbaria is available in the MAPR database and were included in this study. These herbaria are the Herbarium of the Department of Natural and Environmental Resources of Puerto Rico (SJ), the Herbarium of the Biology Department of the University of Puerto Rico at Río Piedras (UPRRP), the Herbarium of the Botanical Garden of the

University of Puerto Rico (UPR), the New York Botanical Garden (NY) and the Herbarium of the Museum of Natural History of the Smithsonian Institute in Washington, D.C. (US).

The records obtained from the MAPR database do not always present accurate descriptions for their localities and in some cases could refer to points outside the forest boundaries. However, most records include descriptive information about their collection localities and in some cases, their geographic coordinates. The process of assigning geographic coordinates to each specimen, or to translate a locality description into a mappable representation of a feature is known as georeferencing (Chapman and Wieckzorek 2006). The specimen information was used to incorporate MAPR data into a geographic information system (GIS) with ArcMap software, following the “point and circle” method described by Chapman and Wieckzorek (2006). According to this method, uncertainties (or errors) for the locality description were calculated for each record to assure the precision of the georeferencing. The data were tabulated by their uncertainty to determine the precision of the collection. They were also tabulated by georeferencing categories corresponding to their locality description to identify the description techniques for most collectors. Records with dubious location information were not georeferenced and those located outside the forest boundaries after georeferencing were not included in the analyses.

The number of records used for species, species-area, infrastructure and temporal bias analysis depended on the nature of the analysis and the maximal uncertainty it can sustain. The records with the lowest uncertainty are best for most of the bias analyses, but they tend to be in low numbers, particularly so for older records, therefore a balance between number of records and uncertainty was attempted for each of the analyses. For the tests that require dividing the study area in a grid of square cells, the records with uncertainties less than to the half size of the cell were used.

Evaluating Bias. The following types of collecting bias were evaluated: taxonomic (only at the species level), spatial (species-area and infrastructure), temporal and environmental bias.

Species Bias. For the plant collection to be representative at the species level there is the assumption that each species was equally collected across the landscape regardless of their distribution and abundance. To assess if this assumption is true, species, subspecies and varieties were treated as separated taxa, according to the nomenclatural criteria used by Axelrod (2011), which is the main nomenclature source used in MAPR. The number of records per species (R_s) was tabulated in a rank abundance curve. Based on the extension of the reserve, its number of species, the number of specimens per species and the personal knowledge of the forest composition, a criterion was established to assess the representativeness of the records in the curve; fewer than four records for the less-collected species and more than 13 records for the more-collected species.

Species-Area Bias. If no bias is present, all species in the collection would be uniformly collected within the area they could occur. To examine the presence of this bias, the GFR map was divided into a grid of square cells of 00°00'17" (approximately 500 m^2). For this analysis, records with uncertainties less than 250 m were used. The number of records per species (R_s) was calculated and the number of grid cells in which each species was collected (G_s). The relationship between R_s and G_s was described with a linear regression. If there is no species–area bias, there would be a high degree of correlation between R_s and G_s, because there would be more species' accessions expected in a larger area. The obtained distribution was compared with a Chi – square test with the expected distribution, if all records were equally collected in relation to the area they occupy in the forest. A P-value less than 0.05 indicate the presence of a species-area bias in the GFR plant collection.

Infrastructure Bias. To test the assumption of collecting uniformly across the landscape, the observed distribution of collection localities for records with uncertainties less than 500 m was analyzed as a function of its arrangement along roads (including trails). Because of the size of the study area, we can visually analyze the collection pattern traced for the GFR by the collectors.

In addition, the study area was divided in a grid with cells of 00°00’35” (approximately 1 km²) and the distance from each cell with one or more collections to the nearest road and to Borinquen Camp was calculated. Borinquen Camp was established by the Civilian Conservation Corps, a state labor program established in 1933 by the United Stated government to help young guys to combat poverty and unemployment during the Great Depression (Monsegur 2009). At present time, Borinquen Camp is the area occupied by the forest office, at the end of paved Road 334, closer to the visitors’ center and was selected to the assessment because its importance as principal reference point in the reserve. The obtained distribution of collection localities from the main road and from Borinquen Camp was compared to a random distribution traced in ArcMap 10 software with a Chi – square test. A P-value less than 0.05 indicate the presence of infrastructure bias in the GFR plant collection.

Temporal Bias. If the collection is not biased the collection effort will be similar over time. To test this assumption, the georeferenced records were tabulated per different variables of time. The distribution of the records was traced over decades to understand the collectors’ effort during historical times. The records distribution were also tabulated over months of the year to clarify the patterns for collectors’ preferences to visit the field and identify which months of the year need more attention than others when sampling plants in the GFR. Using a Chi – square test, the distributions obtained over decades and month were compared to the expected distribution if all records were equally collected through time. If a P-value obtained is less than 0.05, a significant temporal bias is present in the GFR plant collection.

Environmental Bias. In 2002 The Puerto Rico Gap Analysis Program was initiated with a grant to the United States Department of Agriculture (USDA) - International Institute of Tropical Forestry, from the United States Geological Service (USGS). The initial purpose of the project was to compile information of land cover of Puerto Rico to develop species-habitat models and map the distribution of terrestrial invertebrates. The results of this work are presented in Gould *et al.* (2008). The results of the PRGAP land cover classification for the GFR were used as a surrogate of habitat to assess the environmental bias. To test the assumption of collecting uniformly across the environmental gradient, the observed distribution of collection localities for records with uncertainties less than 500 m was analyzed as a function of its arrangement along the different types of land cover. In a grid cell of 00°00'35" (approximately 1 km^2) the number of cells with collection localities was recorded for each land cover type. With a Chi - square test the observed distribution of the collection localities among the different habitats of the forest was compared with the random distribution traced in ArcMap 10 software. A P-value less than 0.05 indicate the presence of the environmental bias in the GFR plant collection.

Results

Georeferencing Records. A total of 2,218 records from MAPR, NY, UPR, UPRRP, US and SJ herbaria were obtained from the MAPR database between December 2009 and April 2013. From those records, 2,188 were located within the GFR boundaries however 1,807 were considered in the bias analyses (Figure 2). Thirty records were not georeferenced, because their location data were too imprecise to identify if they were collected in the GFR. The remaining 381 records were georeferenced but not included in the study because their georeferenced location were outside the GFR boundaries (Table 1).

Of 19 locality types described by Chapman and Wieckzorek (2006), 10 categories were used for the georeferencing process. “Latitude and longitude coordinates” is the most represented category, with 48% of the records in it. These coordinates were calculated from a topographic map, or with a global positioning system (GPS). The second best represented category was “Named place” category (for example, “Guánica Forest”) with 37% of the records, followed by “Path” category (for example, “Cobanas Trail” or “Road 333”) with 9% of the records. The remaining seven categories are represented with less than 2% of the records (Table 2).

According to the locality type described for each record, their uncertainty was calculated to identify the accuracy in accordance with the information they presented in their data labels.A total of 55% of the collection presents uncertainty less than or equal to 500 m, of which 72% have uncertainties less than or equal to 100 m. On the other hand, 10% of the data has intermediate uncertainties (501 to 1,000 m) and a 35% of the records present uncertainties higher than 1,001 m (Table 3).

Species Bias. A total of 88 collectors contributed the 1,807 records for GFR in MAPR database georeferenced in ArcMap. These records accounted for 473 species from 98 families, including three records identified at genus level (Appendix II). They were distributed among nine different life forms following Axelrod’s nomenclature and descriptions (2011). Herbaceous plants was the category best represented with 28.5% (135 spp.) of the species in the collection. The shrubs category, with 18.6 % (88 spp.) of the species, is the second best represented group. Trees were represented by 17.1 % (81 spp.) of the collection while the species that can be shrubs or trees represented 15 % (71 spp.). Vines represented 15 % of the species in the collection (72 spp.). The cacti, ferns, succulents and palms are the less represented life forms with less than 2.1 % of species collected (Table 4).

In order to assess the presence of species bias, a species accumulation curve was traced the 1,807 records georeferenced for the forest (Figure 3). In accordance with the forest area, Monsegur's vascular flora checklist, the number of specimens per species in the MAPR database and after discussed with Dr. J. Danilo Chinea, Dr. Duane Kolterman and Jeanine Velez about our personal knowledge of the flora abundance, it was determined that a less-collected species have four records or less and the more-collected species have 13 or more records collected. We obtained an uneven collecting effort among the species. Only 5% of the species in the forest are more collected while 75 % of them are less-collected and 20% of the species presented an intermediate collection effort (Table 5).

Figure 4 shows a histogram traced in function of the relative frequency of the records collected per species to observe the distribution of the data. A non-normal distribution was obtained, supporting the results of the species accumulation curve, presenting a sub-collection for most of the species and a strong taxonomical bias for the records in the GFR plant collection at MAPR database.

Of the 473 species in the database, 10 were found to occur within the forest boundaries but were not included in Monsegur's 2009 checklist for the forest vascular flora (Table 6). Besides, because MAPR used Axelrod's 2011 name classification system, synonymy and other corrections made for the records were verified with his publication, where he presents an updated nomenclature for the Puerto Rican flora.

Species–Area Bias. For this assessment, the study area was divided into a grid with cells of 00°00'17", which comprises approximately 500 m² (Figure 5). Hence, only records with uncertainties less than 250 m were used, for a total of 942 records. The resulting data were analyzed by the ecological concept that the number of records collected per species will increase if the collecting area increases, obtaining a R^2 value =0.86 and *y=1.16-0.11* as the equation of the line (Figure 6). When assessing the model with an expected distribution, significant differences were found with a P-value obtained < 0.0001 (Appendix III).

Infrastructure Bias. When analyzing the resulting map composition of the georeferenced plant collection for the species–area bias in Figure 5, most of the records with smaller uncertainties (less than 250 m) can be observed being distributed along the trails and roads that traverse the forest. The forest map was then divided in a grid with cells of 00°00'35", which was approximately 1 km², for the infrastructure bias statistical assessment. For this analysis a total of 983 records with uncertainties less than 500 m were used (Figure 7). The distance from the farthest point in each grid cell to the nearest road (Table 7) and from Borinquen Camp was determined (Table 8). Upon assessing the grid cells with collection localities, 76 cells out of 132 were found to have at least one collection, and 80% of the cells were found within 1,000 m from the nearest road. One-fifth of the localities were more than 1,001 m from the nearest road. If the collection localities were randomly distributed (Figure 8), 110 cells out of 132 were found to have at least one collection. The observed distribution of distances from the grid cells to the nearest road was significantly different from the random distribution when compared with a Chi – square test and obtained a P-value less than 0.0001 (Appendix IV).

For the distance between the collected grid cells and the Borinquen Camp, the pattern was very different. A total of 38 cells out of 66 were found to have at least one collection with 79 % of grid cells farther than 2,001 m from the Camp. If the collection localities were randomly distributed (Figure 8), 55 cells were found to have at least one collection with 78% of the grid cells found farther than 2,001 m from Borinquen Camp. After comparing the observed distribution of the distances from the grid cells to the Borinquen Camp with a Chi – square test, significant differences were found with a P-value =0.0217 (Appendix V).

Temporal Bias. To assess the distribution of records through decades, a total of 1,799 records were used. Eight records were not used in the assessment because the collection year was not available in MAPR database. Records collected during 1886 – 1909 were grouped in one category, and identified as the first decade assessed. Table 9 shows that the collection effort was initiated at the end of the 19th Century and slowly increased during the first half of the 20th Century, with a 6% of the collection effort during the 1940's. The database presents a progressively increase since the 1980's with a 13.7% of the effort, 23% during the 1990's and the biggest effort during the first decade of the 21th century with 43% of the records collected (Figure 9). The observed distribution was significantly different from the expected distribution when assessed with the Chi – square test obtaining a P-value less than 0.0001 (Appendix VI).

When assessing the collection effort through the months of the year, 9 records were not used in the assessment due to missing collection date in MAPR database for a total of 1,798 records assessed. The records were divided in the months more collected (more than 10% of the collection effort), intermediate (5 to 9%) and less collected (less than 4% of the effort). Figure 10 shows that the month with more collections was October, with 16 % of the records collected, followed by September (12 % of the collection effort), November, January and June (with 11 % of the effort each month). The months with an intermediate collection effort were December, July, August, April and May, with fluctuations between 5 to 9 % of the total records collected. The less visited months are February and March with only 3 % of the effort in each month (Table 10). The observed distribution was significantly different from the expected distribution when compared with a Chi – square test and obtained a P-value less than 0.0001 (Appendix VII).

Environmental Bias. Figure 11 shows the original classification for the PRGAP project with 20 categories (Gould et al. 2008). To assess the environmental bias with statistical analysis, the original classification was re-classified into seven categories (Table 11). Five categories (dwarf forest, fine to coarse beach, mangrove, rock and cliff and salt water) were not included, because they did not present the minimum quantity of data required to run a Chi – square test.

The grid with cells of 00°00’35” traced for the infrastructure analysis was used to count the cells with at least one collection locality for 983 records with uncertainties less than 500 m. The best represented category is the closed forest on limestone with 47% of the points visited by collectors (Table 12). The second best represented category is the open forest on limestone with 17% of the effort. The remaining categories are represented with less than 10%; open areas like grassland and cacti with 8% each one, followed by non-forest wetland and open forest non-calcareous with 7% each, and for last, the closed forest non-calcareous with 6% of the collection localities. Figure 12 show the random distribution traced in ArcMap 10 to assess the bias in a 1 km^2 scale. After comparing the observed distribution of the collection localities with the random distribution of collection points across the environmental gradient with a Chi – square test, significant differences were found for both distributions with a P-value =0.0161 (Appendix VIII).

Discussion

The GFR now possesses a plant inventory available worldwide through the MAPR database in BRAHMS and a georeferenced database. The most represented categories of locality types for the georeferencing process are those that provide more precision in their location and at the same time, smaller uncertainties, like the latitude and longitude coordinates, collected with a GPS or with a topographic map, and the exact description of their localities. The increases in these categories demonstrate that collectors are being more careful when documenting the exact location of the specimen.

Because of how the collection localities were documented, more than half of the collection (54%) has a precision of 500 m or less. This provides researchers the benefit of using these data in analysis at a local scale to achieve the best management and conservation techniques for the organisms that inhabit the reserve, for example, identify the number of herbarium specimens needed to assesses threatened species (Rivers et al. 2011. However, the records with uncertainties higher than 500 m are not wasteful data. They can be used in projects of a larger scale in the best effort to improve the Island's species conservation; for example, mapping plant diversity for the Guánica municipality or for the Island, in the same way that Schmidt *et al*. (2005).mapped plant diversity in Burkina Faso, Africa. Almost half of the records with uncertainties higher than 1,000 m (315 out of 626 records) are historical specimens dated before 1970 and despite the higher uncertainties they can be useful to assess the regeneration dynamics in a secondary dry forest. This data was collected when the equipment was simpler and not as accurate as the modern technological equipment. In addition, many of these specimens were collected during the first years of the GFR establishment when the reserve boundaries were not clearly identifiable by the collectors. On this basis, in the future it is expected an increase in the collection accuracy due to the availability of high technology equipment that supports the field work.

The GFR has a wide diversity of species from different life forms present in its flora. As discussed by Agosto-Diaz (2008) and Molina-Colón and Lugo (2006) the plant diversity in the GFR can be explained by several factors, such as the wide diversity of habitats, elevation range, the variable topography and the effect of the airborne salt, however, the variable that best describes the species composition is the land-use history; being mature areas those which present higher biodiversity, dominated by native species with the inclusion of some introduced species that should be considered as part of the GFR flora, for example, *Agave sisalana*, *Leucaena leucocephala*, *Megathyrsus maximum*, *Prosopis juliflora* and *Swietenia mahogani* (Chinea 1990, Wadsworth 1990).

The fact that herbs are the best represented life form category (28.5%) is resulting on the numerous open areas present, primarily on limestone substrate. However, Agosto-Diaz (2008) related the species richness with time for canopy recovery. Table 4 shows that the categories of trees and shrubs are represented a total of 50% of the collected specimens when grouped as one category of woody plants. After ca. 100 years of recovery in the GFR, individuals from these two categories are found in forest areas under young or mature secondary succession. As in Hiijmans *et al* (2000) the distribution of the number of records per species was highly biased, showing collectors preference for particular species. Most of the more-collected species, like *Bourreria virgata*, *Coccoloba diversifolia* and *Thouinia portoricensis* are native species widely distributed in the forest which increases the likelihood of their collection (Appendix II). However, as it happened in Hijmans *et al.* (2000) some of the species in the collection are less-collected because are rare or less common. Rare species, such as *Randia portoricensis* and *Zephyranthes proctorii* not neccesarily are sub-collected because they have fewer records in the database, they could be well represenrted according to their population size in the reserve Other species are difficult to collect, like the members of the Cactaceae, because of their large, succulent spiny stems and branches, and the Bromeliaceae, with a morphological structure very difficult to collect. However many of the less-collected species are very common within the reserve boundaries, for example: *Adelia ricinella*, *Gouania lupuloides*, *Gyminda latifolia*, *Hibiscus* sp. and *Randia aculeata*, which, as compared in Schmidt-Lebuhn *et al*. (2013) may be less-collected for not having any specific morphological or phenological characteristic or status that do not represent any particular interest to the collectors.

Of the ten species confirmed by Monsegur (2009) (Appendix I) and documented as possibly extirpated, only three were georeferenced within the GFR forest boundaries and their status within the reserve should be studied further. *Allophylus racemosus* and *Cordia laevigata* are rare, native species documented for more mesic areas on the northern part of the GFR. However, while *Cordia laevigata* can be found in coastal lowlands through the Island, *Allophylus racemosus* prefers mesic habitats at the north of the Central Cordillera (Axelrod 2011). *Ipomoea carnea* subsp. *fistulosa* is an introduced species that should also be assessed as extirpated because it was historically planted by former inhabitants on highly disturbed areas of the forest and apparently did not persist during the forest regeneration. The records for the remaining seven species were not included in the analysis because their locality description did not occur within the forest boundaries. Another ten species were not included in Monsegur's checklist even though they are in the MAPR database as part of the reserve, also need revision (Table 6). Is not very clear that these species not included in the checklist belong to the reserve, and this information should be assessed.

Our results were very similar to those obtained by Hijmans *et. al* (2000) in Bolivia. As expected, the number of records collected in the GFR per species increased with the increasing species distribution area, even though the increase was not proportional for all the species in the forest, as seen in Figure 6 for *Guapira domingensis.* This species was more-collected in the reserve with 15 records, respectively, but they were only collected in five cells. In addition, not all the species that were more-collected according to the species abundance curve are useful for projects at local scale. The three most represented species in the abundance curve (Figure 3), *Thouinia portoricensis*, *Coccoloba diversifolia*, and *Bourreria virgata* have only four, nine, and six records, respectively, with locality descriptions more accurate and uncertainties less than 250 m. As compared with Hijmans *et al.* (2000) we obtained a strong relationship between the two variables, number of records per species and the number of grid cells where species were collected; therefore, there is a species–area bias. These patterns may affect the real distribution of the species and influence the description of biodiversity dynamics in the GFR.

In Kadmon *et al.* (2004) the frequency of plant observations near roads and trails was consistently greater than the expected from a spatially random distribution, and it was most pronounced at distances from 500 m to 2,000 but also at larger scales. Our results were similar to those obtained by Kadmon *et al.* (2004). When tracing the distribution for the GFR records with uncertainties less than 500 m. the comparison between the observed distribution and the expected distribution shows significant differences of collected cells across the roads in the East and West Units of the GFR, with few points far from roads and trails, demonstrating the infrastructure bias (Figure 7). It was most pronounced at distances within 1,000 m from roads, but it was significant at greater distances. Regardless the size of the forest, collectors tend to collect closer to the trails and the roads for reasons of time, security and resource availability. Also, the forest has many trails that cross its terrain and three different main roads that cross the reserve; two roads that cross the East Unit and one road that crosses the West Unit, giving different ways of access to different areas of the forest. The abundance of trails also permits botanists to take their cars close to the trail entrances and walk easily to the farthest zones of the forest and could explain why most part of the grid cells with collection localities are more than 2,000 m from the Borinquen Camp (Figure 7). However, due to the time of recovery of the GFR, ca. 100 years, most of the forest terrains are under a mature or young stage of secondary succession (Agosto-Diaz 2008, Molina-Colón and Lugo 2006) which makes the exploration within the wooded area rather somewhat accessible, and this may explain why we have the areas far from the trails without precise collections showing a spatial bias.

Similar to what Peterson *et al*. (1998) and Soberón *et al*. (2000) documented, the collection effort for GFR was not uniformly distributed through time, showing a temporal bias. Figure 9 display the distribution of collecting effort over time and reflects the pattern of the collecting activity in the Island. The intensity of the collection effort is related with collector's preferences and availability (Hijmans *et. al* 2000, Schmidt-Lebuhn *et al*. 2013, Soberón *et al*. 2000). In the GFR the effort increases when the collectors were available. The peak in collecting effort in the GFR began during the 1940's with the work of different botanists, some of them related to the university and others probably related to US military and political interests although there is no written evidence of this assumption. In the 1950's the effort decay and boost again during the1960's, with the arrival of Dr. Liogier, who worked documenting the Puerto Rican flora. But Dr. Liogier move to the Dominican republic in the 1970's, and it is not until the1980's, that Dr. Acevedo-Rodriguez came to the Island as part of his doctoral preparation, and Dr. Breckon as a faculty member of the University of Puerto Rico at Mayaguez and their collecting activities had a high impact on the GFR plant collection. After the 1980's the plant collection activities increase and in the 1990's Dr. Franklin Axelrod adds his effort and work in the GFR as part of the Biology department of the University of Puerto Rico at Río Piedras.

However, the biggest effort was attributed to the work of the botanist Omar Monsegur with this thesis project documenting the vascular flora of the GFR during the first decade of the 21st Century (2009), contributing with a 43% of the collection effort in this decade. Without this enormous work, the collection effort for this decade would have been minimal and similar to the effort during the 1950's.

When the collection effort was related to the month of the year it was identified an increase during the rainy season in fall (Figure 10). This is the period when the forest productivity and the growth of reproductive structures increase because rainfall levels approaches or exceed the potential evapotranspiration (Murphy and Lugo 1990). The peak during the summer, in July can be explained by the low rainy season in late spring.

Even all the land cover types present in the GFR are represented in the vascular plant specimen collection available in the MAPR database; there is the presence of the bias, because there are significant differences in the number of records that should be collected for each habitat type if the sampling were planned with a random structure (Table 12).

Hortal *et al*. (2007) assessed the environmental bias as function of spatial bias and obtained a poor predictive performance when modeling species richness in Tenerife. On the contrary, other authors have addressed the environmental bias and the results of their projects were not highly influenced by the environmental bias as it happened with the spatial bias or other kind of bias (Hortal *et. al* 2007, Kadmon *et. al* 2004, Loiselle *et al.* 2008). In our case, the environmental bias can be understood as the under-representation of some habitats. This can be related to the collectors' interest in select areas with specific characteristics which led to an aggregated survey pattern directed to areas with greater values of species as possible (Satre and Lobo 2009). Species richness in the GFR can be explained by several factors such as water availability, elevation, substrate and disturbance history. Also, the wide diversity of habitats can be the result of the variable topography and the land use history (Monsegur 2009, Murphy and Lugo 1990, Wadsworth 1990). As previously discussed the species diversity in a recovered mature forest was higher than those areas of the GFR in a pre-conversion stage (Chinea 1990; Molina-Colón and Lugo 2006; Wadsworth 1990). Figure11 shows the land cover classification for the forest and the bigger areas are the forest on limestone, which is the predominant soil type of the forest (Murphy and Lugo 1990). Table 7 confirms the pattern of preferences by collectors drive by special characteristics, special interests or higher species richness. The forest on limestone is the biggest land cover of the forest and is predominant in a mature stage because of its ca 100 years of recuperation. According to Molina-Colón and Lugo (2006) this is the environment that will present higher biodiversity. The next categories in order of representation are the open areas with grassland and cacti that are dominant in different parts of the forest, with the herbaceous plants as the dominant life form in the species composition (when assessing trees, shrubs and shrub/tree as separate life forms). Areas of high importance due to its unique composition (mangrove, beach, rock cliffs and the dwarf forest) needs concentrated effort to increase their representation on the MAPR database.

Environmental bias is assessed primarily by climatic patterns (Lioselle et al. 2008, Kadmon et. al 2004) but in a tropical island this is not a variable that best describes differences in habitat in the GFR. Agosto-Diaz (2008) and Molina-Colón and Lugo (2006) demonstrates that the most important variable determining species composition is the land use history and the land cover is the best surrogate for this assessment. Even though all the land cover types present collection localities, the presence of a significant bias toward a limited number of habitats (dwarf forest, beaches, mangroves, rock cliffs and salt water) shows that additional surveys should be conducted to cover the entire spectrum of environmental conditions and the geographic extent (Hortal *et. al* 2007). This will improve sampling designs used in conservation planning and decision making.

Conclusions

The most important findings of this study are highlighted in the following points:

1. There is a georeferenced database in ArcMap 10 for the GFR with 1,807 records from 473 species distributed in 98 families obtained from MAPR database. These data were georeferenced to the greatest possible accuracy resulting in more than 50% of the data useful to develop conservation and management projects on a local scale.
2. The GFR georeferenced database presents a significant bias by species, species-area, infrastructure, time and environment.
3. Most part of the plant collection is less-represented with abundance of species with only one record in the MAPR database.
4. According to the ecological concepts, collections per species will increase if the collected areas increase. Some species did not comply with this assumption warning to be very careful when interpreting species distribution based on museum data.
5. The distribution of records that are useful for projects at local scale area spatially biased across roads and trails. This analysis reveals the need to move away from the road and explore areas to obtain better results about the composition of the forest.
6. The collection effort increase through the decades following a collecting pattern with a peak during the high rainy season of the year in the GFR. This bias was expected because of the lack of historical information, but should be keep in focus to avoid decay the effort obtained in the past decade
7. There is a representation of records collected in all the diverse habitats in the GFR, but the representation is biased towards the larger types of land cover. Even so, increasing the collections in the sub-represented regions will improve the ecological dynamics understanding in hostile environments.

This plant records collection stored in an electronic database is representative of the species composition in the GFR; however it is not representative of the biodiversity. Therefore, it is necessary to address new collection techniques based on existing knowledge to guide future expeditions, inventories and surveys. The West Unit and the east part of the East Unit deserve more attention by collectors, as same as under-represented environments like mangroves, salts, wetlands and beaches. The analysis of the GFR plant collection points out another dimension in which the collection needs improvement; filling these gaps through continued scientific collecting would represent an important step in completing our understanding in dry forest plants of the world.

Because of the importance of this project is necessary to report these results to the scientific community to increase the knowledge about the advantages of having a good, unbiased sampling method and the benefits that the Guánica Forest presents as a focus for plant ecology projects.

Strategic Collection Plan

To increase the quality of the herbarium data available for the GFR and achieve the final goal of this project, some ideas are suggested to be taking into account by the collectors during their expeditions.

1. Focus new collecting efforts in the less visited areas and during the first months of the year. These collecting trips could be planned with the MAPR herbarium to include them as part of their activities and include the participation of undergraduate students in the collection trips. This not only will decrease the sampling bias found in the collection and increase its usefulness in scientific research, but will also train new students, interested in Botany, in the best techniques to obtain records of high quality. Also, this will lead collectors to explore areas that can be suitable for species different than those found in more accessible sites and could be a way to find information about those species that could be extirpated or find new species for the collection.

2. Give special attention to the less–collected species, to increase their abundance in records maintained in the herbarium, including the endangered species. This would strengthen the knowledge about the species' actual distribution in the forest and will promote biodiversity assessment at different landscape levels.

3. Record bias during botanical expeditions and surveys to minimize bias adopting a controlled, systematic and repeatable method. Documentation of what was done, where, when and whom can help with interpretation of the data, but these details are rarely documented (Rich and Woodruff 1992). The addition of photography, field journals and new technological devices will help to improve the quality of the data obtained and decrease the resulting bias in the collections.

Other activities are suggested to keep track of this work and extrapolate it to other areas on the Island:

1. Repeat the assessment process on a basis that could be estimated in five to ten years. This will give a continuous status of which areas should be of priority attention.
2. Maintain a continuous and active georeferencing process for the GFR flora training the collectors, at least for the material maintained at MAPR to include material collected after the analysis.
3. Integrate the georeferenced database to MAPR database.
4. Generate alternative forms of documentation, such as photographs and journals of botanical obseravtions (supported by BRAHMS) for those species difficult to collect.
5. Begin the georeferencing of records of other areas of high importance on the Island, for example: reserves such as El Yunque, Caja de Muertos and the Karst Belt, and adding eventually all collections, including those located in private lands.
6. Assess the bias in collections for other areas of conservation importance and investigate how the collecting patterns compare among them.
7. Update the MAPR website to promote worldwide scientific research on plants in the dry forest and other habitats in the Island.
8. New publications focused on the scientific community, such as "Acta Científica" to update the scientific information available about the research in the GFR.
9. Feedback the MAPR database with the existing georeferenced database for the GFR.
10. Promote a process to standardize collection of plants that can be used worldwide among herbaria, universities and any research institution that works with plants assessment.

Literature Cited

Agosto-Díaz, R. E. 2008. Human and Environmental Factors Explaining the Structural and Compositional Variability in a Sub-tropical Dry Forest. M. S. Thesis, University of Puerto Rico, Mayagüez Campus. 55 pp.

Álvarez–Ruíz, M., V. L. Santiago and A. Puente. 1990. El Bosque de Guánica como Recurso de Investigación Científica. Acta Científica 4 (1-3): 3-14.

Axelrod, F. S. 2011. A Systematic Vademecum to the Vascular Plants of Puerto Rico. Botanical Research Institute of Texas. 420 pp.

Biodiversity Monitoring & Assessment Project. 2005. http://www.biomapegypt.org/index1.html. April, 2015

Brown, A. H. D., and D. R. Marshall. 1995. A Basic Sampling Strategy: Theory and Practice. In L. Guarino, V. Ramantha Rao, and R. Reid (eds). Collecting plant genetic diversity. Technical Guidelines, pp. 75-91. CAB International, Wallingford, United Kingdom.

Canals-Mora, M. E. 1990. El Futuro del Bosque de Guánica como una Unidad Efectiva de Conservación. Acta Científica 4 (1-3): 109-112.

Canary Islands Government. 1999. Sistema de Información Medioambiental de Canarias: Banco de Datos de Biodiversidad de Canarias. http://www.gobiernodecanarias.org/medioambiente/piac/

Chapman, A. D. and J. Wieczorek. 2006. Guide to Best Practices for Georeferencing. Copenhagen: Global Biodiversity Information Facility. The Regents of the University of California. 80 pp.

Chinea, J. D. 1990. Árboles Introducidos a la Reserva de Guánica. Acta Científica 4 (1-3): 51-59.

Department of Natural and Environmental Resources, Commonwealth of Puerto Rico. 1981. The Master Plan for the Commonwealth Forests of Puerto Rico. Area of Planning and Evaluation Resources. Division of Forest Planning.

Ewel, J. J. and J. L. Whitmore. 1973. The Ecological Life Zones of Puerto Rico and the US Virgin Islands. USDA Forest Service, Institute of Tropical Forestry.

Finnish Museum of Natural History. Atlas Florae Europaeae. University of Helsinki. www.luomus.fi/english/botany/afe/index.htm January, 2014.

Global Biodiversity Information Facility. 2001. http://www.gbif.org/. April, 2015

Gould, W. A., C. Alarcón, B. Fevold, M. E. Jiménez, S. Martinuzzi, G. Potts, M. Quiñones, M. Solórzano and E. Ventosa. 2008. The Puerto Rico Gap Analysis Project. Volume 1: Land cover, vertebrate species distributions, and land stewardship. Department of Agriculture, Forest Service, International Institute of Tropical Forestry. Río Piedras, PR. 165 pp.

Graham, C. H., S. Ferrier, F. Huettman, C. Moritz and A. Townsend Peterson. 2004. New Developments in Museum-based Informatics and Applications in Biodiversity Analysis. Trends in Ecology and Evolution 19 (9): 497-503.

Hijmans, R. J., K. A. Garrett, Z. Huamán, D. P. Zhang, M. Schreuder and M. Bonierbale. 2000. Assessing the Geographic Representativeness of Genebank Collections: the Case of Bolivian Wild Potatoes. Conservation Biology 14: 1755-1765.

Hortal, J., J. M. Lobo, and A. Jímenez-Valverde. 2007. Limitations of Biodiversity Databases: Case Study on Seed-Plant Diversity in Tenerife, Canary Islands. Conservation Biology 21: 3: 853-863.

Janzen, D. H. 1988. Tropical Dry Forest: The Most Endangered Major Tropical Ecosystem. In Wilson, E. O. (ed.). Biodiversity, pp. 130-137. National Academic Press, Washington, D.C.

Kadmon, R., O. Farber, and A. Danin. 2004. Effect of Roadside Bias on the Accuracy of Predictive Maps Produced by Bioclimatic Models. Ecological Applications 14: 401-413.

Loiselle, B. A., P. M. Jorgensen, T. Consiglio, I. Jiménez, J. G. Blake, L. G. Lohmann and O. M. Montiel. 2008. Predicting Species Distributions from Herbarium Collections: Does Climate Bias in Collection Sampling Influence Model Outcomes? Journal of Biogeography 35: 105-116.

Lugo, A. E., O. Ramos, S. Molina, F. N. Scatena and L. L.Vélez. 1996. A Fifty-three Year Record of Land Use Change in the Guánica Forest Biosphere Reserve and its Vicinity. Río Piedras, U.S. Dept. of Agriculture, Forest Service, International Institute of Tropical Forestry with Fundación Puertorriqueña de Conservación.

Lugo, A. E. 1990. El Movimiento Conservacionista y las Reservas Biosfericas: Alternativas y Soluciones al Problema de Desarrollo en Puerto Rico. Acta Científica 4 (1-3): 97-107.

Margules, C. R. and R. L. Pressey. 2000. Systematic Conservation Planning. Nature 405: 243—253.

Merriam-Webster Incorporated 2015. Merriam-Webster Online Dictionary http://www.merriam-webster.com/dictionary/bias/ March 30, 2015.

Molina-Colón, S. and A. E. Lugo. 2006. Recovery of a Subtropical Dry Forest After Abandonment of Different Land Uses. Biotropica 38 (3) 354 – 364.

Monsegur, O. 2009. Vascular Flora of The Guánica Dry Forest, Puerto Rico. M. S. Thesis, University of Puerto Rico, Mayagüez Campus. 215 pp.

Murphy, P. G. and A. E. Lugo. 1990. Dry Forest of the Tropics and Subtropics: Guánica Forest in Context. Acta Científica 4 (1-3): 15-24

Newbold, T. 2010. Applications and Limitations of Museum Data for Conservation and Ecology, with Particular Attention to Species Distribution Models. Progress in Physical Geography 34 (1): 3-22

Pearson, R. G. 2008. Species' Distribution Modelling for Conservation Educators and Practitioners. Center for Biodiversity and Conservation and Department of Herpetology. American Museum of Natural History.

Peterson, A. T., A. G. Navarro-Siguenza and H. Benítez-Díaz. 1998. The Need for Continued Scientific Collecting: A Geographic Analysis of Mexican Bird Specimens. Ibis 140: 288-294.

Puerto Rico Gap Analysis Program. http://prgap.org/about/prgap/. April, 2015.

Rich, T. C. G. and E. R. Woodruff. 1992. Recording Bias in Botanical Surveys. Watsonia 19: 73-95.

Rivers, M. C., L. Taylor, N. A. Brummitt, T. R. Meagher, D. L. Roberts and E. N. Lughadha. 2011. How Many Herbarium Specimens are Needed to Detect Threatened Species? Biological Conservation 144: 2541-2547.

Sastre, P and J. M. Lobo. 2009. Taxonomist Survey Bias and the Unveiling of Biodiversity Patterns. Biological Conservation 142: 462-467.

Schmidt, M., H. Kreft, A. Thiomniano and G. Zizka. 2005. Herbarium Collections and Field Data-based Plant Diversity Maps for Burkina Faso. Diversity and Distributions 11: 509-516

Schmidt-Lebuhn, A. N., N. J. Knerr and M. Kessler. 2013. Non-Geographic Collecting Biases in Herbarium Specimens of Australian Daisies (Asteraceae). Biodivers Conservation 22: 905-919.

Soberón, J., and A. T. Peterson. 2004. Biodiversity Informatics: Managing and Applying Primary Biodiversity Data. Philosophical Transactions of the Royal Society of London B 359: 689-698.

Soberón, J. M., J. B. Llorente and L. Oñate. 2000. The Use of Specimen-Label Databases for Conservation Purposes: an Example using Mexican Papilionid and Pierid butterflies. Biodiversity and Conservation 9: 1441 – 1466.

Sousa-Baena, M. S., L. Couto Garcia and A. Townsend Peterson. 2013. Completeness of Digital Accesible Knowledge of the Plants of Brazil and Priorities for Survey and Inventory. Diversity and Distribution: 1-13.

Stockwell, D. R. B., and A. T. Peterson. 2002. Controlling Bias in Biodiversity Data. Pages 537-546 in J. M. Scott, P. J. Heglund, J. B. Haufler, M. Morrison, M. G. Raphael, W. B. Wall, and E. Samson, editors. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Covelo, California.

The United Nations, Educational, Scientific and Cultural Organization. 2014. Man and Biosphere Programme. http://www.unesco.org/mabdb/br/brdir/directory/biores.asp?mode=genandcode=USA+35. April, 2011.

University of Oxford, Department of Plant Sciences. 1985. BRAHMS Online: Botanical Research and Herbarium Management System. http://herbaria.plants.ox.ac.uk/bol/. April, 2015.

University of Oxford, Department of Plant Sciences. 1985. BRAHMS Online: UPR Mayagüez Herbarium-MAPR. Department of Biology, University of Puerto Rico. http://herbaria.plants.ox.ac.uk/bol/mapr. April, 2015

Urban, J. 1911. Symbolae Antillanae: seu Fundamenta Florae Indiae Occidentalis. Florae Portoricensis. 772 pp.

Wadsworth, F. W. Plantaciones Forestales en el Bosque Estatal de Guánica. Acta Científica 4 (1-3): 61-68.

Williams, P. H., C. R. Margules and D. W. Hilbert. 2002. Data requirements and Data Sources for Biodiversity Priority Area Selection. J. Biosci 27 (2): 327-228.

Tables

Table 1. Number of "Guánica" identified records in the MAPR database.

Reason for exclusion	*Number of records*
Dubious location	30
Outside GFR boundary	381
Inside GFR boundary	1,807

Table 2. Distribution of GFR georeferenced records according to locality types.

Locality type[1]	*Number of records*	*Percentage (%)*
Lat/Long Coordinates	860	48
Named place	668	37
Path	162	9
Near a feature	38	2
Offset along a path	25	1
Offset distance	20	1
Between two features	11	1
Offset direction	9	<1
Offset at a heading	9	<1
Between two paths	5	<1
Total of records	*1807*	

[1] ***According to Chapman and Wieckzoreck 2006***

Table 3. Distribution of georeferenced records according to magnitude of uncertainties.

Uncertainty range (m)	*Number of records*	*Percentage (%)*
0 – 100	720	40
101 – 250	203	11
251 – 500	74	4
501 – 1000	184	10
>1001	626	35
Total of records	*1807*	

Table 4. Distribution of species according to life form. See Appendix I for definitions.

Life form	*Number of species*	*Percentage (%)*
Herbs	135	28.5
Shrubs	88	18.6
Trees	81	17.1
Shrubs / Trees	71	15.0
Vines	72	15.0
Cacti	10	2.1
Ferns	7	1.5
Succulents	7	1.5
Palms	2	<1
Total of species	*473*	

Table 5. Distribution of georeferenced records according to number of records per species *(Rs)*.

Records per species	*Number of species*	*Percentage (%)*
1 – 4	353	75
5 – 12	95	20
>13	25	5
Total of species	*473*	

Table 6. Species in MAPR database for GFR not listed in Monsegur's checklist.

Species	*Family*
Ambrosia peruviana	Asteraceae
Cassuarina cunninghamiana	Casuarinaceae
Dysphania	Amaranthaceae
Euphorbia oerstediana	Euphorbiaceae
Guadua fruticosa	Poaceae
Lagenaria siceraria	Cucurbitaceae
Mitracarpus portoricensis	Rubiaceae
Phoradendron berteroana	Santalaceae
Porophyllum ruderale	Asteraceae
Rhynchosia pyramidilis	Fabaceae – Faboideae
Stenandrium droseroides	Acanthaceae

Table 7. Distribution of cells with collection localities to assess the infrastructure bias in relation to the nearest roads.

Distance to nearest road (m)	*Observed number of cells with collection locality*	*Percentage(%)*
<1,000	61	91
>1,001m	15	9
Total of grid cells counted	*76*	

Table 8. Distribution of cells with collection localities to assess the infrastructure bias in relation to the distance from Borinquen Camp.

Distance to Borinquen Camp (m)	*Observed number of cells with collection locality*	*Percentage (%)*
<2,000	8	21
>2,001m	30	79
Total of grid cells counted	*38*	

Table 9. Distribution of records per decade for the GFR for the temporal assessment.

Decade	*Number of records*	*Percentage (%)*
1886	7	0.4
1910 – 1919	5	0.2
1920 – 1929	5	0.2
1930 – 1939	23	1.3
1940 – 1949	108	6.1
1950 – 1959	65	3.7
1960 – 1969	137	7.7
1970 – 1979	17	1.1
1980 – 1989	245	13.7
1990 – 1999	413	23
2000 – 2009	773	43
Total of records counted	*1798*	

Table 10. Distribution of records per month for the GFR for the temporal assessment.

Decade	*Number of records*	*Percentage (%)*
January	199	11
February	57	3
March	62	3
April	89	5
May	81	5
June	196	11
July	139	8
August	118	7
September	207	12
October	289	16
November	203	11
December	158	9
Total of records	*1798*	

Table 11. List of categories of land cover classification for the GFR resulting for Gould *et al.'s* (2000) PRGAP project and the resulting categories for the statistical analysis to assess the environmental bias.

PRGAP Land Cover Categories	*Statistical Analysis Categories*
Mature secondary lowland dry limestone deciduous forest	Closed forest on limestone (Mature and Young)
Mature secondary lowland dry limestone evergreen forest	Closed forest on limestone (Mature and Young)
Mature secondary lowland dry limestone semideciduous forest	Closed forest on limestone (Mature and Young)
Young secondary lowland dry limestone deciduous forest	Closed forest on limestone (Mature and Young)
Young secondary dry lowland limestone semideciduous forest	Closed forest on limestone (Mature and Young)
Lowland dry limestone shrubland	Open forest on limestone (Shrub and Woodland)
Lowland dry limestone wood & shrubland	Open forest on limestone (Shrub and Woodland)
Mature secondary lowland dry alluvial semideciduous forest	Closed forest non-calcareous (Mature and Young)
Mature secondary non-calcareous dry semideciduous forest	Closed forest non-calcareous (Mature and Young)
Young secondary lowland alluvial semideciduous forest	Closed forest non-calcareous (Mature and Young)
Lowland alluvial shrub and woodland	Open forest non-calcareous (Shrubland and Woodland)
Salts & mudflats	Non forest wetland (Salts, mudflats, seasonal flood)
Seasonally flood herbaceous saline wetland	Non forest wetland (Salts, mudflats, seasonal flood)
Dry cactus & grassland	Cacti
Dry grassland & pasture	Grassland
Dwarf forest	Not used
Fine to coarsy sandy beach, sand and gravel	Not used
Mangrove and shrubland	Not used
Rock and cliff shelves	Not used
Salt water	Not used

Table 12. Distribution of grid cells per land cover category for the environmental assessment.

Land cover category	*Observed number of cells with collection locality*	*Percentage*
Closed forest on limestone	40	47
Open forest on limestone	15	17
Closed forest non-calcareous	6	7
Open forest non-calcareous	5	6
Non forest wetland	5	7
Grassland	8	8
Cacti	7	8
Total of grid cells counted	*86*	

Figures

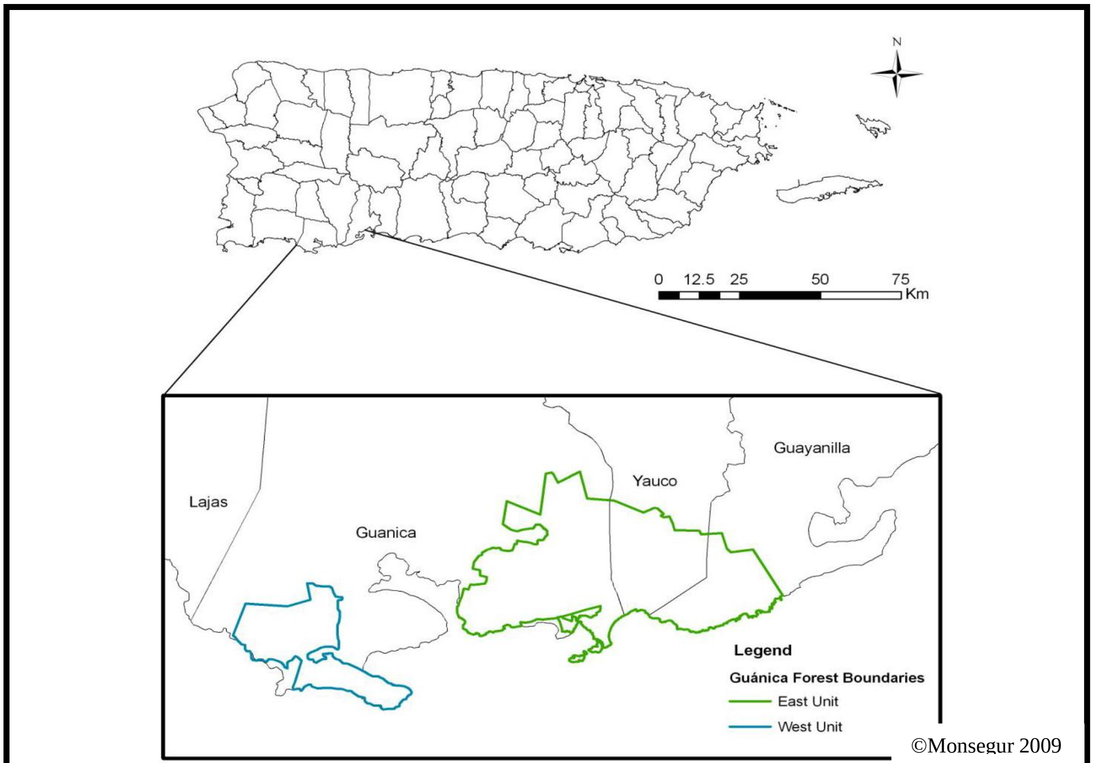

Figure 1: Map of the location of the Guánica Forest Reserve, including the West and East Units.

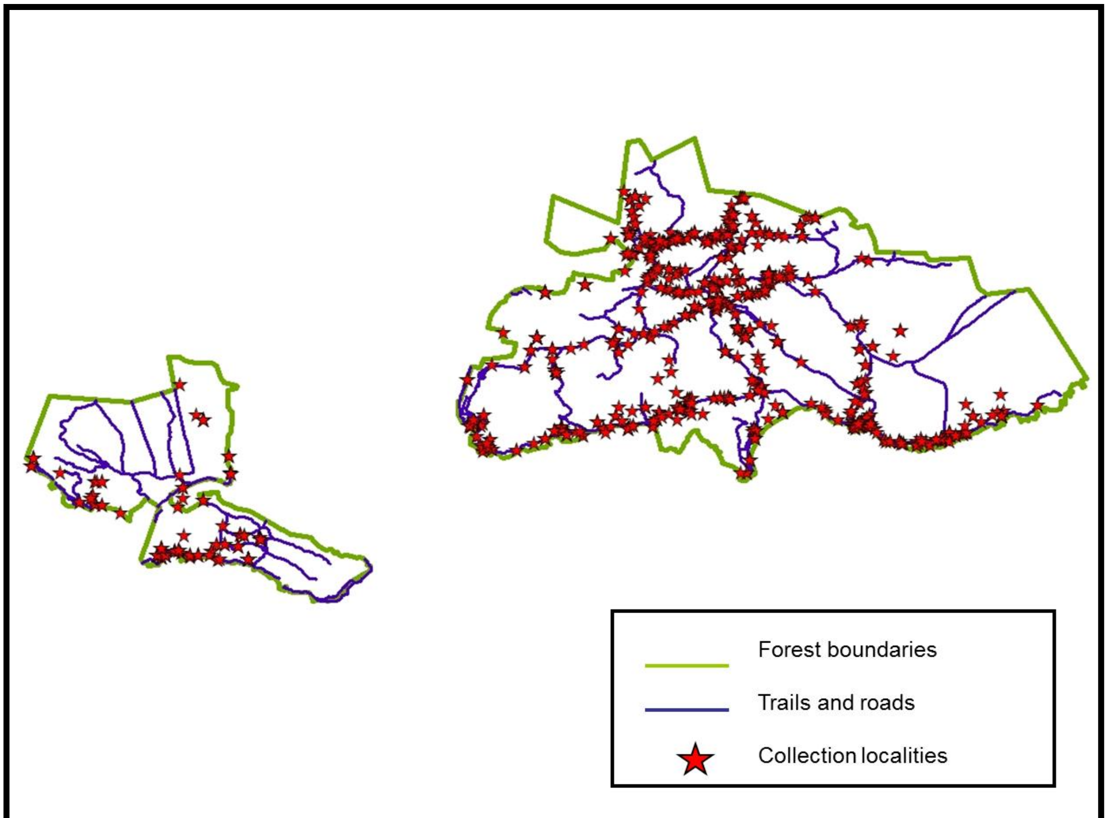

Figure 2: Map of the collection localities for the GFR plant records available in the MAPR database. The green lines represent the forest new boundaries, the blue lines are the roads and trails and the red stars represent the collection localities for 1,807 georeferenced records.

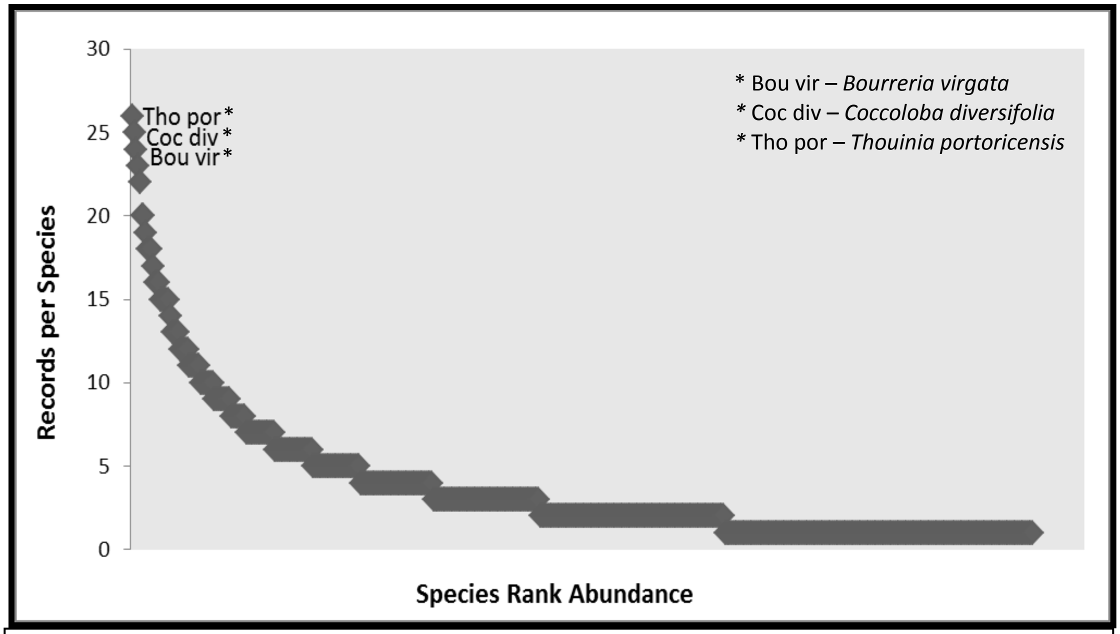

Figure 3: Species Rank Abundance Curve for the GFR collected data available at MAPR database.

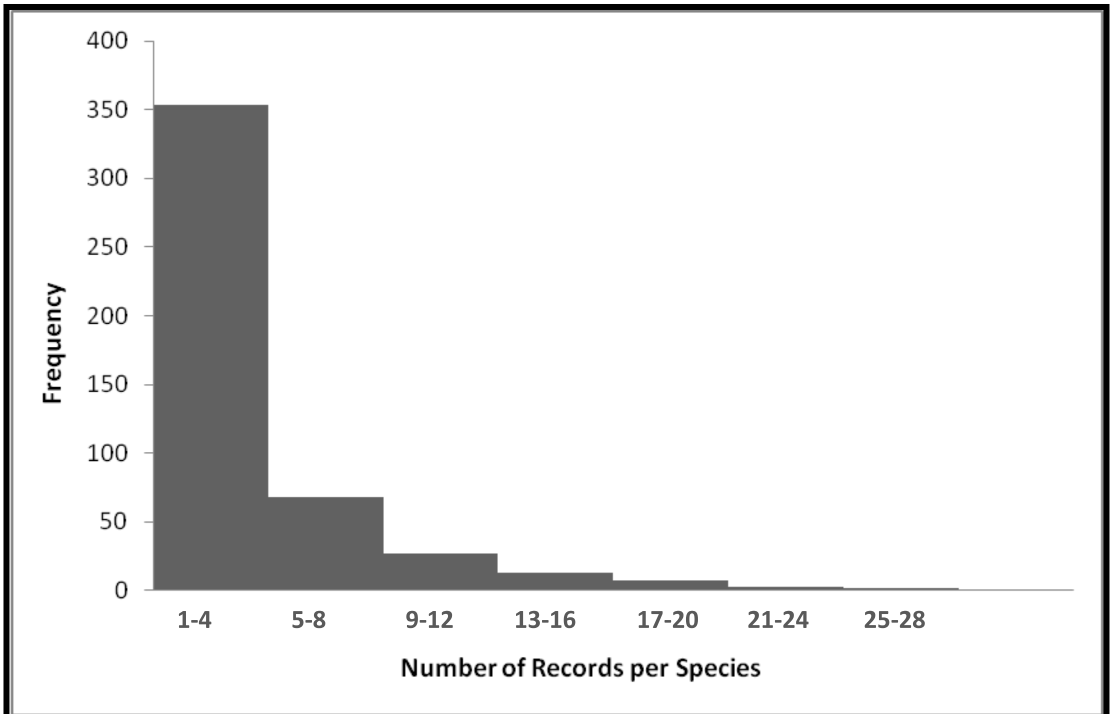

Figure 4: Distribution of records per Species for the GFR data available at MAPR database.

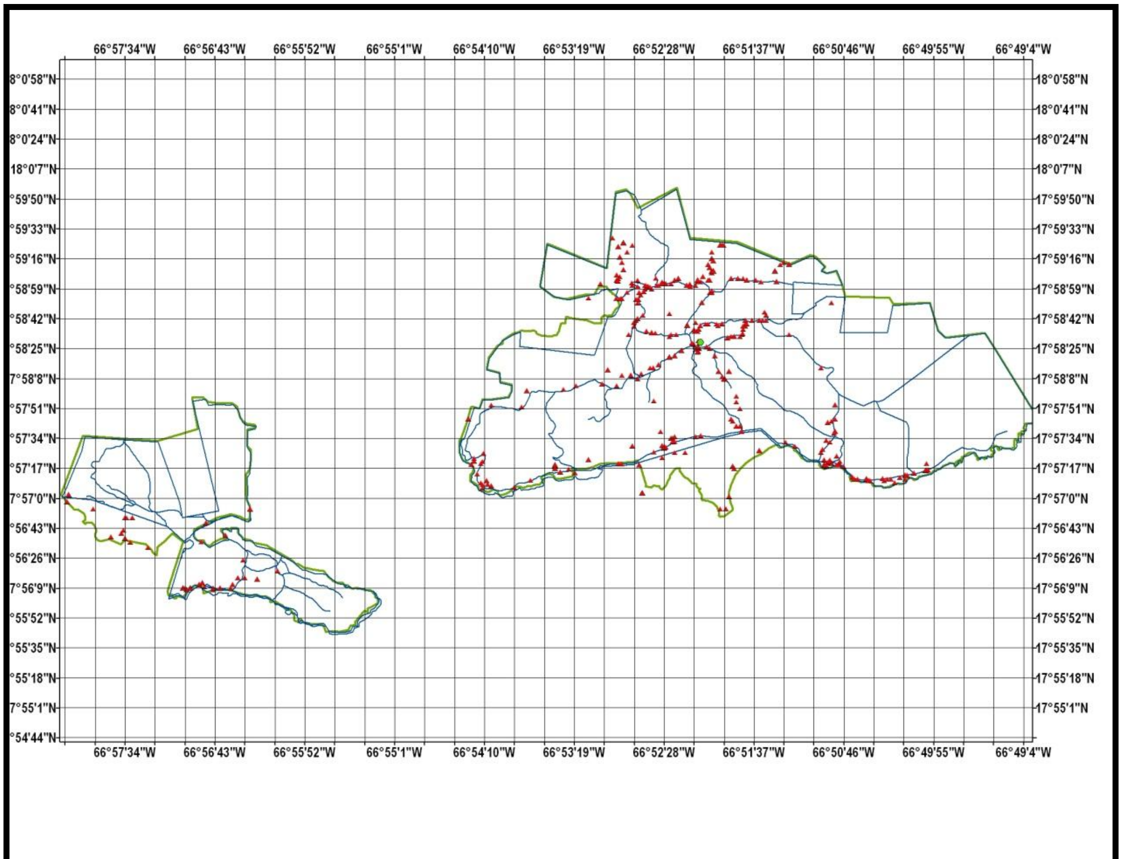

Figure 5: Grid with cells of 00°00’17” to assess spatial bias at the 500 m² scale. The red stars are the records with uncertainties less than to 250 m. The green dot represents Borinquen Camp. The blue lines are the roads and trails across the forest.

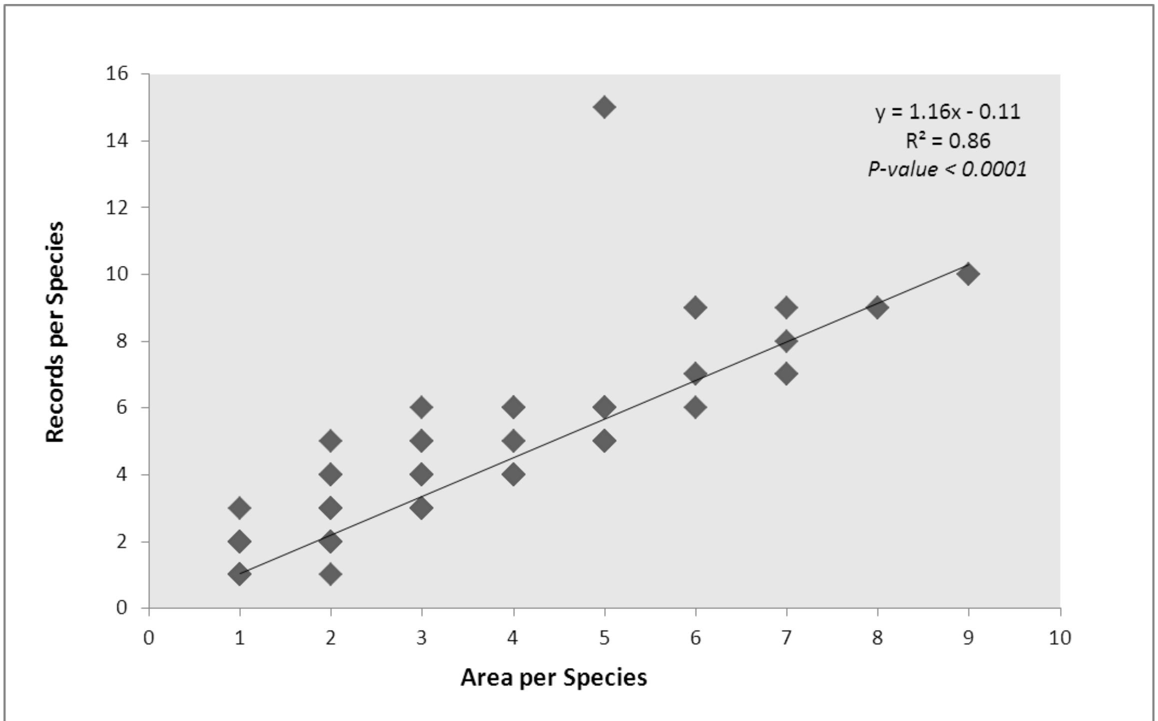

Figure 6. Linear regression traced for the collection effort for each species as a function of the area they occupy in the GFR.

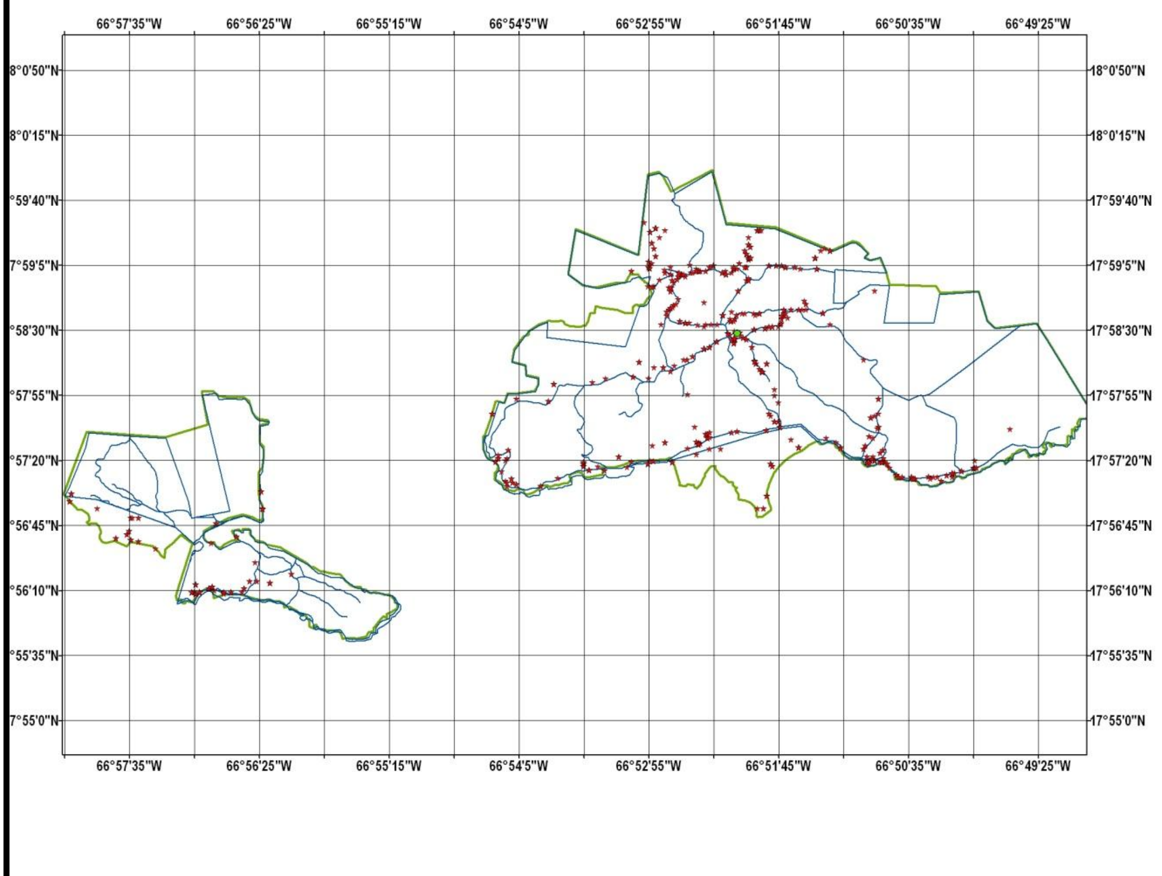

Figure 7: Grid with cells of 00°00'35" to assess spatial bias at the 1 km² scale. The red stars are the records with uncertainties less than 500 m used to examine the spatial bias. The blue lines are the roads and trails across the forest.

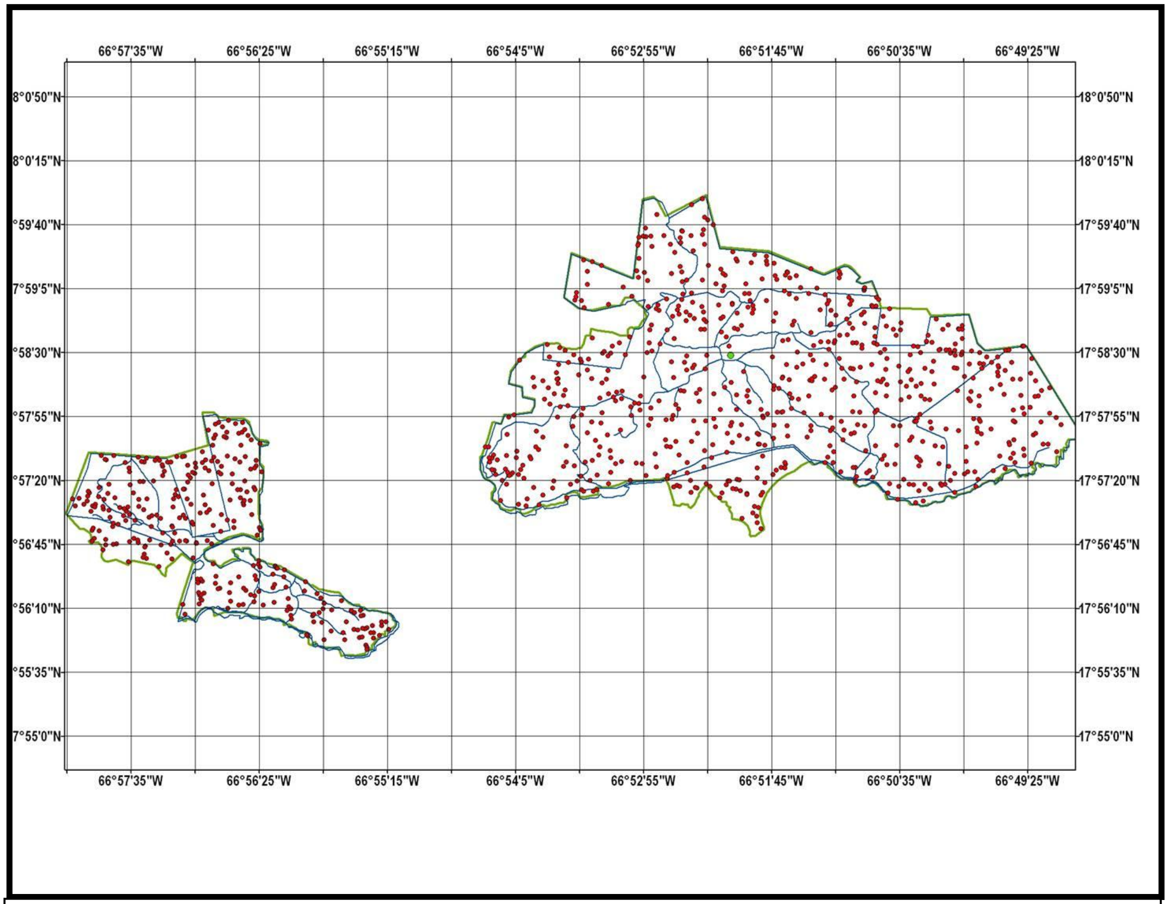

Figure 8: Grid cell of 00°00’35” with 1,000 random points generated to assess the infrastructure bias at 1 km² scale in the GFR plant collection available in MAPR database.

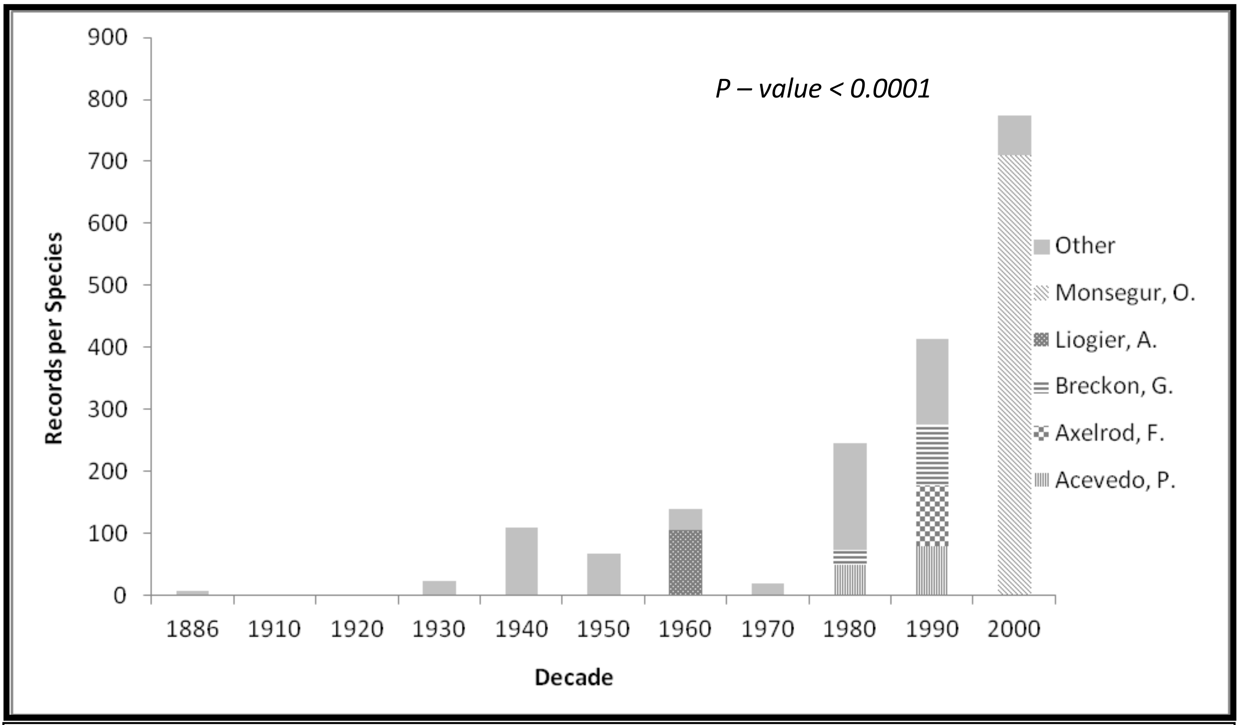

Figure 9. Distribution of the collecting effort in the GFR through the decades.

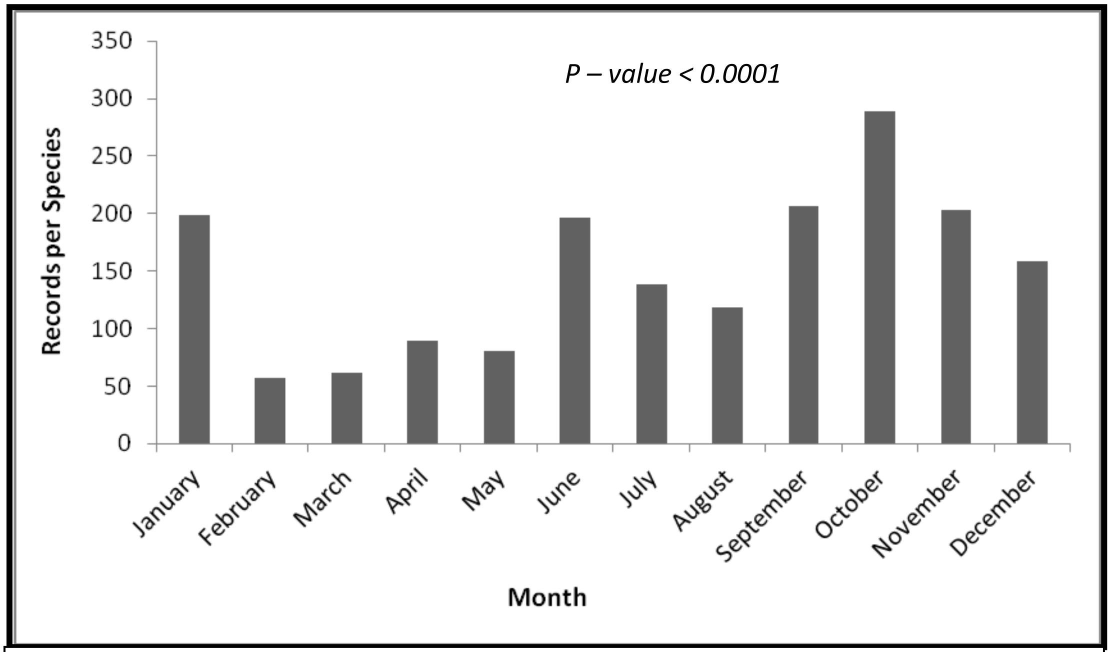

Figure 10. Distribution of the collecting effort in the GFR through the months.

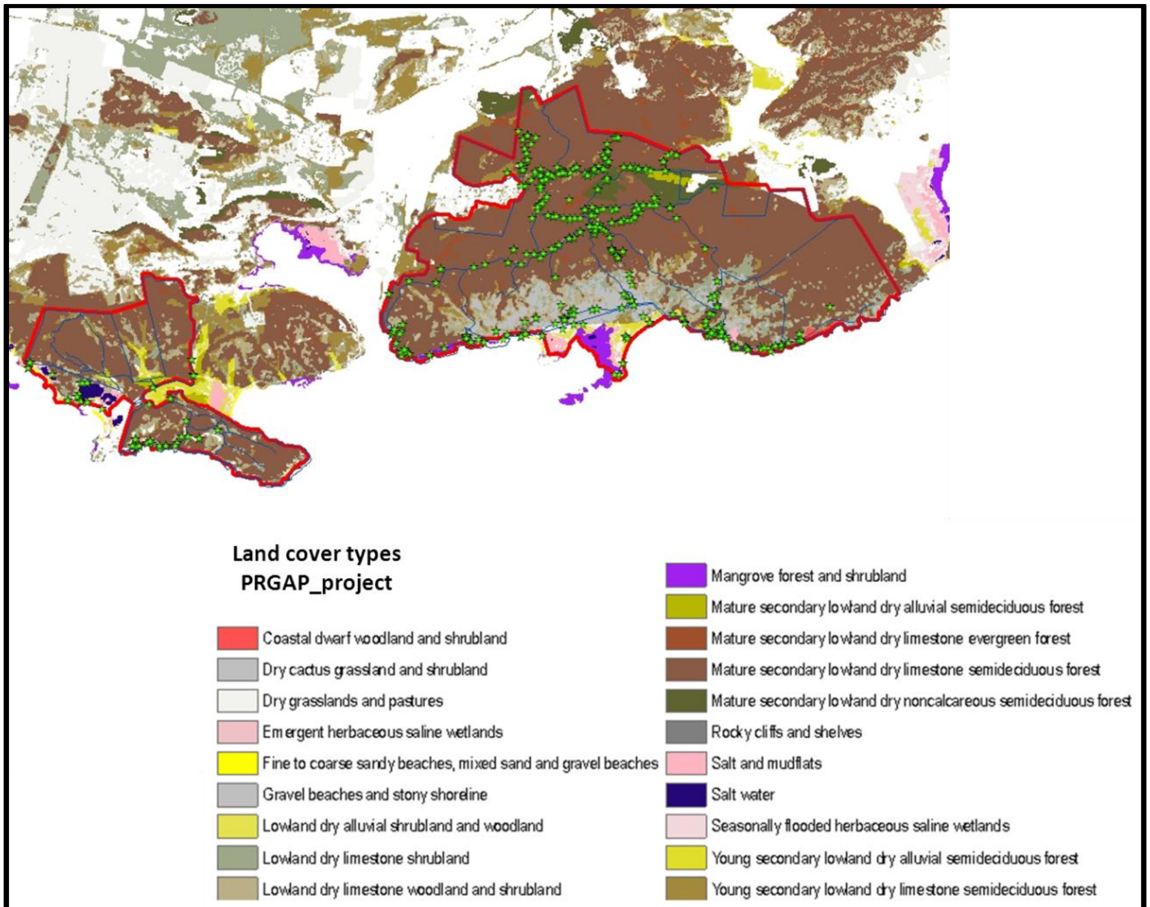

Figure 11: Map of the collection localities for the GFR plant records available in the MAPR database. The green stars are the collection localities for records with uncertainties less than 500 m used to examine the spatial bias. The red lines are the forest boundaries. The blue lines are the roads and trails.

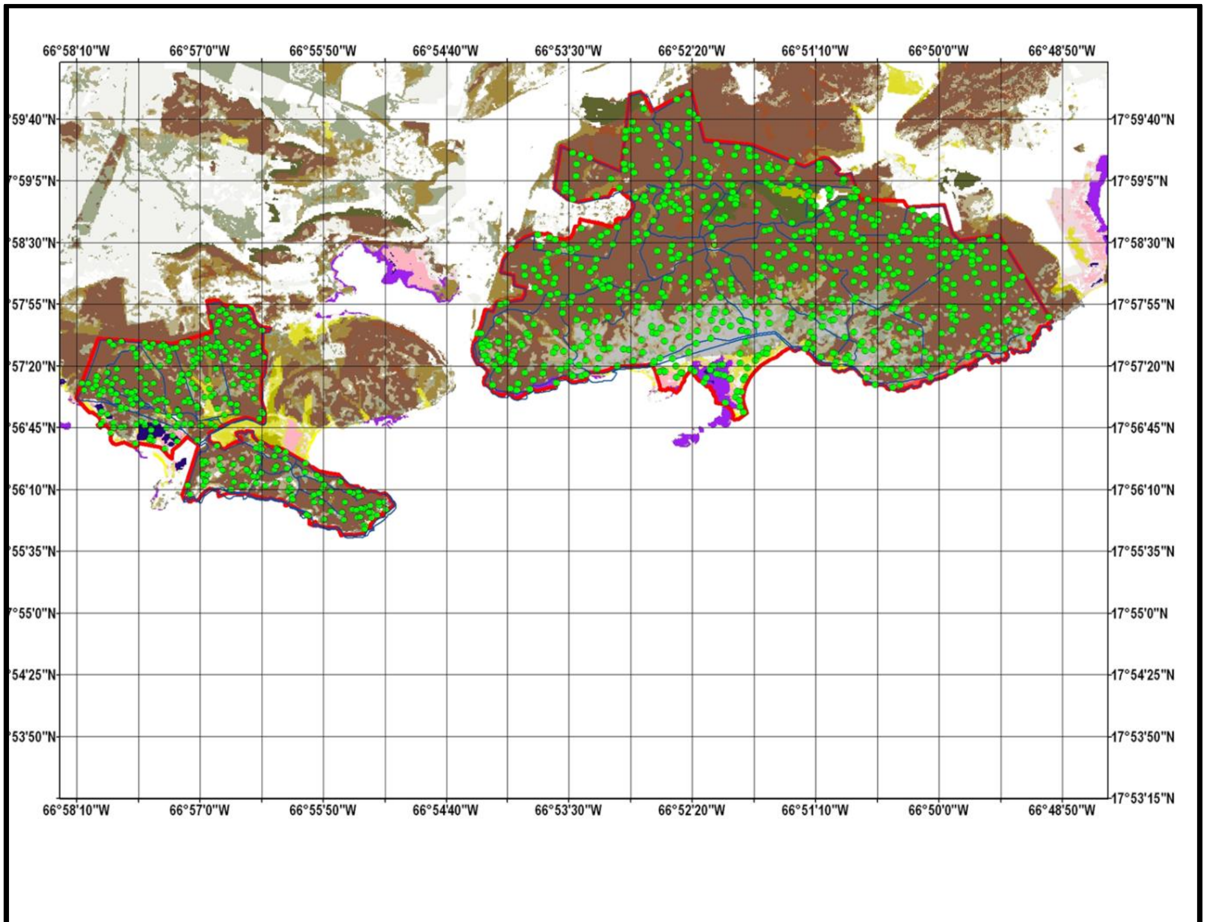

Figure 12: Grid cell of 00°00’35” with 1,000 random points generated through a land cover classification to assess the environmental bias at 1 km² scale in the GFR plant collection available in MAPR database.

Appendices

Appendix I. Glossary

Confirmed Species – Species whose identification was verified with a specimen collected within the forest boundary or with a specimen preserved in herbaria, and whose locality could be determined to occur within the forest boundary.

Epiphyte – A plant that derives its moisture and nutrients from the air and rain and grows on another plant.

Fern – Any of a large class of flowerless spore – producing vascular plants; any of a division of homosporous plants possessing roots, stems and leaf-like fronds.

Growth Habit – A way of classifying plants alternative to the ordinary species-genus-family scientific classification; based on the stages in the process of growth of their stems.

Herb – A seed-producing annual, biennial or perennial that does not develop persistent woody tissue but dies down at the end of its growing season.

Palm – any member of the family Araceae (or Palmae) of mostly tropical or subtropical monocotyledonous trees, shrubs or vines with a simple stem and a terminal crown of large pinnate or fan–shaped leaves.

Shrub – A low, several-stemmed woody plant.

Shrub / Tree – An individual that can present either of the two growth habits in any stage of the life cycle, or depending on the habitat.

Succulent – A plant that has fleshy tissue that conserves moisture.

Tree – A woody perennial plant usually having a single elongate main stem (trunk) generally with few or no branches on its lower part.

Unconfirmed Species – Species whose identification was not verified with a specimen collected within the forest boundary or with a dried specimen preserved in herbaria. Hence, its locality was not able to be confirmed to occur within the forest boundary.

Vine – A plant whose stem requires support and which climbs by tendrils or twining or creeps along the ground.

Appendix II. List of species in MAPR database with its number of records, the life form it presents and its status on the Island.

Family	*Species*	*Number of Records*	*Growth Form*	*Status*
Acanthaceae	*Avicennia germinans*	3	S / T	N
Acanthaceae	*Justicia periplocifolia*	2	H	N
Acanthaceae	*Justicia sessilis*	4	H	N
Acanthaceae	*Oplonia spinosa*	2	S	N
Acanthaceae	*Ruellia tuberosa*	4	H	N
Acanthaceae	*Stenandrium droseroides*	1	H	N
Acanthaceae	*Stenandrium tuberosum*	2	H	N
Agavaceae	*Agave missionum*	1	Succ	E
Agavaceae	*Agave sisalana*	3	Succ	I
Agavaceae	*Furcraea tuberosa*	1	Succ	N
Agavaceae	*Yucca aloifolia*	1	Succ	I
Aizoaceae	*Sesuvium portulacastrum*	7	Succ	N
Amaranthaceae	*Achyranthes aspera* var. *aspera*	2	H	I
Amaranthaceae	*Amaranthus crassipes*	1	H	N
Amaranthaceae	*Atriplex cristata*	4	H	N
Amaranthaceae	*Celosia nitida*	6	H	N
Amaranthaceae	*Dysphania* sp.	1	H	N
Amaranthaceae	*Lithophilla muscoides*	4	H	N
Amaryllidaceae	*Zephyranthes proctorii*	2	H	E
Anacardiaceae	*Comocladia dodonaea*	5	T	N
Anacardiaceae	*Spondias dulcis*	1	T	I
Annonaceae	*Annona reticulata*	2	T	I
Annonaceae	*Oxandra lanceolata*	1	T	N
Apocynaceae	*Calotropis procera*	3	S	I
Apocynaceae	*Cryptostegia madagascariensis*	5	V	I
Apocynaceae	*Echites agglutinatus*	6	V	N
Apocynaceae	*Marsdenia woodburyana*	5	V	E
Apocynaceae	*Matelea maritima*	2	V	N
Apocynaceae	*Metastelma decipiens*	1	V	N
Apocynaceae	*Metastelma lineare*	3	V	N
Apocynaceae	*Metastelma monense*	1	V	E
Apocynaceae	*Metastelma parviflorum*	3	V	N
Apocynaceae	*Pentalinon luteum*	2	V	N
Apocynaceae	*Plumeria alba*	12	T	N
Apocynaceae	*Rauvolfia nitida*	4	S / T	N
Apocynaceae	*Rauvolfia vidris*	10	S / T	N
Apocynaceae	*Thevetia peruviana*	3	S / T	I
Araceae	*Anthurium crenatum*	1	H	N
Arecaceae	*Sabal causiarum*	1	P	N
Arecaceae	*Thrinax morrisii*	3	P	N
Asphodelaceae	*Aloe vera*	2	Succ	I
Aspleniaceae	*Asplenium heterochorum*	1	F	N
Asteraceae	*Ambrosia peruviana*	1	H	N
Asteraceae	*Chaptalia dentata*	2	H	N
Asteraceae	*Chromolaena odorata*	1	S	N
Asteraceae	*Chromolaena sinuata*	2	S	N
Asteraceae	*Cyanthillium cinereum*	2	H	I

Asteraceae	*Launaea intybacea*	2	H	N
Asteraceae	*Pectis linifolia*	3	H	N
Asteraceae	*Pluchea carolinensis*	1	H	N
Asteraceae	*Porophyllum ruderale*	1	H	N
Asteraceae	*Wedelia calycina*	10	H	N
Bataceae	*Batis maritima*	3	Succ	N
Bignoniaceae	*Crescentia linearifolia*	1	S / T	N
Bignoniaceae	*Distictis lactiflora*	6	V	N
Bignoniaceae	*Macfadyena unguis-cati*	3	V	N
Bignoniaceae	*Mansoa alliacea*	1	V	N
Bignoniaceae	*Tabebuia heterophylla*	13	T	N
Bignoniaceae	*Tecoma stans*	1	S / T	I
Boraginaceae	*Bourreria baccata*	19	S / T	N
Boraginaceae	*Bourreria virgata*	24	S / T	N
Boraginaceae	*Cordia collococca*	1	T	N
Boraginaceae	*Cordia laevigata*	2	S / T	N
Boraginaceae	*Cordia obliqua*	1	T	I
Boraginaceae	*Cordia rickseckeri*	4	T	E
Boraginaceae	*Heliotropium angiospermum*	4	S	N
Boraginaceae	*Heliotropium curassavicum*	2	H	N
Boraginaceae	*Rochefortia acanthophora*	2	S / T	N
Boraginaceae	*Rochefortia spinosa*	1	S / T	N
Boraginaceae	*Tournefortia scabra*	2	V	N
Boraginaceae	*Tournefortia volubilis*	10	V	N
Boraginaceae	*Varronia bullata* subsp. *humilis*	7	S	N
Boraginaceae	*Varronia curassavica*	3	S	N
Boraginaceae	*Varronia rupicola*	5	S	E
Brassicaceae	*Cakile lanceolata*	3	H	N
Bromeliaceae	*Tillandsia fasciculata*	1	H	N
Bromeliaceae	*Tillandsia flexuosa*	1	H	N
Bromeliaceae	*Tillandsia recurvata*	1	H	N
Bromeliaceae	*Tillandsia setacea*	2	H	N
Bromeliaceae	*Tillandsia utriculata*	1	H	N
Bromeliaceae	*Tillandsia variabilis*	3	H	N
Burseraceae	*Bursera simaruba*	16	T	N
Cactaceae	*Hylocereus trigonus*	2	C	N
Cactaceae	*Leptocereus quadricostatus*	1	C	E
Cactaceae	*Melocactus intortus*	1	C	N
Cactaceae	*Opuntia cochenillifera*	1	C	I
Cactaceae	*Opuntia repens*	1	C	E
Cactaceae	*Opuntia rubescens*	2	C	N
Cactaceae	*Opuntia stricta*	1	C	N
Cactaceae	*Pilosocereus royenii*	4	C	N
Cactaceae	*Stenocereus fimbriatus*	2	C	N
Cactaceae	*Stenocereus peruvianus*	1	C	N
Canellaceae	*Canella winterana*	10	S / T	N
Capparaceae	*Cynophalla amplissima*	1	T	N
Capparaceae	*Cynophalla flexuosa*	10	S / T	N
Capparaceae	*Cynophalla hastata*	5	S	N
Capparaceae	*Quadrella cynophallophora*	3	S / T	N

Capparaceae	*Quadrella indica*	4	S / T	N
Casuarinaceae	*Casuarina cunninghamiana*	1	T	N
Casuarinaceae	*Casuarina equisetifolia*	4	T	I
Celastraceae	*Crossopetalum rhacoma*	13	T	N
Celastraceae	*Elaeodendron xylocarpum*	9	T	N
Celastraceae	*Gyminda latifolia*	4	T	N
Celastraceae	*Schaefferia frutescens*	2	S	N
Cleomaceae	*Arivela viscosa*	1	H	I
Cleomaceae	*Cleome stenophylla*	1	H	N
Clusiaceae	*Clusia gundlachii*	2	V	E
Clusiaceae	*Clusia rosea*	2	T	N
Combretaceae	*Bucida buceras*	6	T	N
Combretaceae	*Conocarpus erectus*	5	T	N
Combretaceae	*Laguncularia racemosa*	2	T	N
Combretaceae	*Terminalia catappa*	1	T	I
Commelinaceae	*Callisia fragrans*	1	H	I
Commelinaceae	*Commelina erecta*	3	H	N
Convolvulaceae	*Convolvulus nodiflorus*	3	V	N
Convolvulaceae	*Cuscuta americana*	1	V	N
Convolvulaceae	*Cuscuta globulosa*	1	V	N
Convolvulaceae	*Cuscuta umbellata*	1	V	N
Convolvulaceae	*Evolvulus convolvuloides*	1	H	N
Convolvulaceae	*Evolvulus sericeus*	3	H	N
Convolvulaceae	*Ipomoea carnea* var. *fistulosa*	1	V	I
Convolvulaceae	*Ipomoea hederifolia*	1	V	N
Convolvulaceae	*Ipomoea nil*	1	V	N
Convolvulaceae	*Ipomoea pes-caprae*	1	V	N
Convolvulaceae	*Ipomoea setifera*	1	V	N
Convolvulaceae	*Ipomoea steudellii*	11	V	E
Convolvulaceae	*Ipomoea triloba*	2	V	N
Convolvulaceae	*Ipomoea violacea*	5	V	N
Convolvulaceae	*Jacquemontia cayensis*	3	V	N
Convolvulaceae	*Jacquemontia cumanensis*	15	V	N
Convolvulaceae	*Jacquemontia havanensis*	2	V	N
Convolvulaceae	*Jacquemontia penthanthos*	1	V	N
Convolvulaceae	*Jacquemontia solanifolia*	3	V	N
Convolvulaceae	*Merremia aegyptia*	2	V	N
Convolvulaceae	*Merremia dissecta*	2	V	N
Crassulaceae	*Kalanchoe daigremontiana*	4	H	I
Cucurbitaceae	*Doyerea emetocathartica*	3	V	N
Cucurbitaceae	*Lagenaria siceraria*	1	V	N
Cymodoceaceae	*Halodule wrightii*	1	H	N
Cyperaceae	*Abildgaardia ovata*	2	H	N
Cyperaceae	*Bulbostylis curassavica*	3	H	N
Cyperaceae	*Bulbostylis pauciflora*	1	H	N
Cyperaceae	*Cyperus brunneus*	3	H	N
Cyperaceae	*Cyperus elegans*	1	H	N
Cyperaceae	*Cyperus unifolius*	5	H	N
Cyperaceae	*Eleocharis geniculata*	1	H	N
Cyperaceae	*Fimbristylis cymosa*	4	H	N

Cyperaceae	*Scleria lithosperma*	8	H	N
Dioscoreaceae	*Rajania cordata*	2	V	N
Erythroxylaceae	*Erythroxylum aerolatum*	20	S / T	N
Erythroxylaceae	*Erythroxylum brevipes*	4	S / T	N
Erythroxylaceae	*Erythroxylum rotundifolium*	12	S / T	N
Euphorbiaceae	*Acalypha portoricensis*	1	S	E
Euphorbiaceae	*Adelia ricinella*	1	S / T	N
Euphorbiaceae	*Argythamnia candicans*	3	S	N
Euphorbiaceae	*Argythamnia fasciculata*	6	S	N
Euphorbiaceae	*Bernardia dichotoma*	3	S	N
Euphorbiaceae	*Cnidoscolus aconitifolius*	1	S / T	I
Euphorbiaceae	*Croton betulinus*	6	S	N
Euphorbiaceae	*Croton discolor*	14	S	N
Euphorbiaceae	*Croton flavens* var. *rigidus*	8	S	E
Euphorbiaceae	*Croton glabellus*	7	S	N
Euphorbiaceae	*Croton lucidus*	7	S	N
Euphorbiaceae	*Dalechampia scandens*	2	V	N
Euphorbiaceae	*Euphorbia articulata*	2	S	N
Euphorbiaceae	*Euphorbia berteroana*	3	H	N
Euphorbiaceae	*Euphorbia cowellii*	4	H	N
Euphorbiaceae	*Euphorbia cyathophora*	1	H	N
Euphorbiaceae	*Euphorbia hirta*	2	H	N
Euphorbiaceae	*Euphorbia lactea*	1	S	N
Euphorbiaceae	*Euphorbia mesembrianthemifolia*	1	S	N
Euphorbiaceae	*Euphorbia oerstediana*	1	H	N
Euphorbiaceae	*Euphorbia petiolaris*	1	S / T	N
Euphorbiaceae	*Euphorbia serpens*	1	H	N
Euphorbiaceae	*Euphorbia thymifolia*	1	H	N
Euphorbiaceae	*Euphorbia tithymaloides* subsp. *tithymaloides*	1	S	I
Euphorbiaceae	*Euphorbia turpinii*	10	H	N
Euphorbiaceae	*Gymnanthes lucida*	11	S / T	N
Euphorbiaceae	*Hippomane mancinella*	2	T	N
Euphorbiaceae	*Jatropha gossypiifolia*	1	S	N
Euphorbiaceae	*Jatropha hernandiifolia*	6	S	N
Euphorbiaceae	*Tragia volubilis*	2	V	N
Fabaceae-Caesalpinioideae	*Caesalpinia pulcherrima*	1	S / T	I
Fabaceae-Caesalpinioideae	*Chamaecrista lineata* var. *brachyloba*	5	S	N
Fabaceae-Caesalpinioideae	*Delonix regia*	1	T	I
Fabaceae-Caesalpinioideae	*Guilandina bonduc*	1	V	N
Fabaceae-Caesalpinioideae	*Haematoxylum campechianum*	1	T	I
Fabaceae-Caesalpinioideae	*Parkinsonia aculeata*	2	S / T	I
Fabaceae-Caesalpinioideae	*Senna polyphylla* var. *polyphylla*	6	S / T	E
Fabaceae-Caesalpinioideae	*Senna uniflora*	3	H	N
Fabaceae-Caesalpinioideae	*Stahlia monosperma*	2	T	N
Fabaceae-Caesalpinioideae	*Tamarindus indica*	5	T	I
Fabaceae-Faboideae	*Andira inermis*	2	T	N
Fabaceae-Faboideae	*Canavalia rosea*	1	V	N
Fabaceae-Faboideae	*Centrosema virginianum*	5	V	N
Fabaceae-Faboideae	*Clitoria ternatea*	1	V	I
Fabaceae-Faboideae	*Coursetia caribaea*	1	S	N

Fabaceae-Faboideae	*Crotalaria lotifolia*	1	S	N
Fabaceae-Faboideae	*Dalea carthagenensis* var. *portoricana*	2	S	E
Fabaceae-Faboideae	*Desmodium glabrum*	4	H	N
Fabaceae-Faboideae	*Galactia dubia*	4	V	N
Fabaceae-Faboideae	*Galactia striata*	6	V	N
Fabaceae-Faboideae	*Indigofera spicata*	1	H	I
Fabaceae-Faboideae	*Macropitilium lathyroides*	1	H	N
Fabaceae-Faboideae	*Pictetia aculeata*	8	S / T	E
Fabaceae-Faboideae	*Poitea florida*	3	S / T	E
Fabaceae-Faboideae	*Rhynchosia pyramidalis*	1	H	N
Fabaceae-Faboideae	*Sesbania sericea*	2	S	N
Fabaceae-Faboideae	*Stylosanthes hamata*	5	H	N
Fabaceae-Faboideae	*Tephrosia cinerea*	6	H	N
Fabaceae-Faboideae	*Teramnus labialis*	2	V	N
Fabaceae-Mimosoideae	*Desmanthus virgatus*	2	S	N
Fabaceae-Mimosoideae	*Leucaena leucocephala*	9	S / T	I
Fabaceae-Mimosoideae	*Pithecellobium unguis-cati*	15	S / T	N
Fabaceae-Mimosoideae	*Prosopis juliflora*	7	T	I
Fabaceae-Mimosoideae	*Prosopis pallida*	1	S / T	N
Fabaceae-Mimosoideae	*Senegalia riparia*	2	S / T	N
Fabaceae-Mimosoideae	*Vachellia farnesiana*	2	S	I
Fabaceae-Mimosoideae	*Zapoteca portoricensis*	2	S	N
Gesneriaceae	*Gesneria pedunculosa*	1	S / T	E
Goodeniaceae	*Scaveola plumieri*	1	S	N
Hydrocharitaceae	*Thalassia testudinum*	1	H	N
Icacinaceae	*Ottoschulzia rhodoxylon*	7	T	N
Krameriaceae	*Krameria ixine*	18	S	N
Lamiaceae	*Clerodendrum aculeatum*	1	S	N
Lamiaceae	*Ocimum campechianum*	2	H	N
Lauraceae	*Cassytha filiformis*	1	V	N
Lauraceae	*Licaria parvifolia*	3	T	N
Lauraceae	*Nectandra patens*	2	S / T	N
Loganiaceae	*Spigelia anthelmia*	3	H	N
Loranthaceae	*Dendropemom caribaeus*	1	S	N
Loranthaceae	*Dendropemon purpureus*	3	S	N
Malpighiaceae	*Bunchosia glandulosa*	5	S / T	N
Malpighiaceae	*Byrsonima lucida*	2	S / T	N
Malpighiaceae	*Heteropterys purpurea*	7	V	N
Malpighiaceae	*Stigmaphyllon emarginatum*	9	V	N
Malpighiaceae	*Stigmaphyllon floribundum*	2	V	E
Malvaceae	*Ayenia insulicola*	6	H	N
Malvaceae	*Bastardia viscosa* var. *viscosa*	2	H	N
Malvaceae	*Corchorus aestuans*	1	H	N
Malvaceae	*Corchorus hirsutus*	12	S	N
Malvaceae	*Gossypium barbadense* var. *acuminatum*	1	S	N
Malvaceae	*Gossypium hirsutum* var. *marie-galante*	1	S	I
Malvaceae	*Gossypium hirsutum* x *barbadense*	5	S	I
Malvaceae	*Guazuma ulmifolia*	2	T	N
Malvaceae	*Helicteres jamaicensis*	6	T	N
Malvaceae	*Herrisantia crispa*	3	H	N

Malvaceae	*Hibiscus clypeatus*	2	S	N
Malvaceae	*Hibiscus phoeniceus*	5	S	N
Malvaceae	*Malachra alceifolia*	1	H	N
Malvaceae	*Melochia tomentosa*	6	S	N
Malvaceae	*Pavonia spinifex*	3	S	N
Malvaceae	*Pseudabutilon umbellatum*	2	S	N
Malvaceae	*Sida abutifolia*	4	H	N
Malvaceae	*Sida ciliaris*	2	H	N
Malvaceae	*Thespesia populnea*	2	T	N
Malvaceae	*Waltheria indica*	3	S	N
Meliaceae	*Melia azedarach*	1	T	N
Meliaceae	*Swietenia mahogani*	2	T	I
Meliaceae	*Trichilia hirta*	1	T	N
Meliaceae	*Trichilia pallida*	1	T	N
Meliaceae	*Trichilia triacantha*	9	T	E
Moraceae	*Ficus citrifolia*	9	T	N
Moraceae	*Ficus trigonata*	2	T	N
Moraceae	*Maclura tinctoria*	1	T	N
Moringaceae	*Moringa oleifera*	1	S / T	I
Myrsinaceae	*Ardisia obovata*	1	S	N
Myrtaceae	*Calyptranthes pallens*	4	S / T	N
Myrtaceae	*Eugenia axillaris*	6	S / T	N
Myrtaceae	*Eugenia foetida*	22	S	N
Myrtaceae	*Eugenia ligustrina*	4	T	N
Myrtaceae	*Eugenia monticola*	4	S / T	N
Myrtaceae	*Eugenia rhombea*	15	S / T	N
Myrtaceae	*Eugenia woodburyana*	6	T	E
Myrtaceae	*Mosiera xerophytica*	7	S / T	E
Myrtaceae	*Myrcianthes fragrans*	3	T	N
Myrtaceae	*Myrciaria borinquena*	2	S	E
Nephrolepidaceae	*Nephrolepis brownii*	1	F	I
Nyctaginaceae	*Boerhavia coccinea*	1	H	N
Nyctaginaceae	*Boerhavia diffusa*	3	H	N
Nyctaginaceae	*Boerhavia erecta*	3	H	N
Nyctaginaceae	*Bougainvillea × buttiana*	1	V	I
Nyctaginaceae	*Commicarpus scandens*	5	S	N
Nyctaginaceae	*Guapira discolor*	2	S / T	N
Nyctaginaceae	*Guapira domingensis*	15	S / T	N
Nyctaginaceae	*Guapira fragrans*	2	S / T	N
Nyctaginaceae	*Guapira obtusata*	5	S / T	N
Nyctaginaceae	*Guapira* sp.	1	S / T	N
Nyctaginaceae	*Neea buxifolia*	1	S	E
Nyctaginaceae	*Pisonia aculeata*	1	S / T	N
Nyctaginaceae	*Pisonia albida*	20	T	N
Olacaceae	*Ximenia americana*	6	S / T	N
Oleaceae	*Chionanthus holdridgei*	4	S / T	E
Oleaceae	*Forestiera eggersiana*	1	S / T	N
Oleaceae	*Forestiera segregata*	5	S / T	N
Oleaceae	*Jasminum fluminense*	2	S	I
Orchidaceae	*Dandrophylax porrectus*	3	V	N

Orchidaceae	*Ionopsis utricularioides*	2	V	N
Orchidaceae	*Mesadenus lucayanus*	1	V	N
Orchidaceae	*Oeceoclades maculata*	3	V	I
Orchidaceae	*Psychilis krugii*	10	V	E
Orchidaceae	*Tolumnia variegata*	2	V	N
Orchidaceae	*Vanilla barbellata*	3	V	N
Papaveraceae	*Argemone mexicana*	1	H	N
Passifloraceae	*Passiflora berteroana*	1	V	N
Passifloraceae	*Passiflora bilobata*	4	V	N
Passifloraceae	*Passiflora edulis*	1	V	I
Passifloraceae	*Passiflora suberosa*	5	V	N
Phyllanthaceae	*Flueggea acidoton*	1	S / T	N
Phyllanthaceae	*Phyllanthus epiphyllanthus*	2	S	N
Phyllanthaceae	*Phyllanthus pentaphyllus* subsp. *polycladus*	7	S	N
Phyllanthaceae	*Savia sessiliflora*	4	S	N
Phytolaccaceae	*Petiveria alliacea*	1	H	N
Phytolaccaceae	*Rivina humilis*	2	H	N
Picramniaceae	*Picramnia pentandra*	3	T	N
Piperaceae	*Peperomia humilis*	1	H	N
Piperaceae	*Piper amalago*	2	S	N
Plumbaginaceae	*Plumbago scandens*	1	H	N
Poaceae	*Aristida adscensionis*	4	H	N
Poaceae	*Arundo donax*	1	H	I
Poaceae	*Bothriochloa pertusa*	4	H	I
Poaceae	*Bouteloua juncea*	2	H	N
Poaceae	*Bouteloua repens*	3	H	N
Poaceae	*Cenchrus ciliaris*	3	H	I
Poaceae	*Cenchrus echinatus*	1	H	N
Poaceae	*Cenchrus incertus*	1	H	N
Poaceae	*Chloris barbata*	1	H	N
Poaceae	*Dactyloctenium aegyptium*	2	H	I
Poaceae	*Echinochloa colona*	1	H	N
Poaceae	*Eleusina indica*	1	H	I
Poaceae	*Eragrostis ciliaris* var. *laxa*	1	H	I
Poaceae	*Eragrostis tenella*	1	H	I
Poaceae	*Eriochloa polystachya*	1	H	N
Poaceae	*Guadua fruticosa*	6	H	N
Poaceae	*Heteropogon contortus*	2	H	N
Poaceae	*Lasiacis divaricata*	8	H	N
Poaceae	*Lithachne pauciflora*	2	H	N
Poaceae	*Megathyrsus maximus*	3	H	I
Poaceae	*Melinis repens*	2	H	I
Poaceae	*Pappophorum pappiferum*	3	H	N
Poaceae	*Paspalum laxum*	1	H	N
Poaceae	*Paspalum vaginatum*	1	H	N
Poaceae	*Pharus lappulaceus*	1	H	N
Poaceae	*Setaria pradana*	1	H	N
Poaceae	*Setaria setosa* var. *setosa*	5	H	N
Poaceae	*Sporobolus jacquemontii*	3	H	N
Poaceae	*Sporobolus pyramidatus*	1	H	N

Poaceae	*Tragus berteronianus*	3	H	I
Poaceae	*Uniola virgata*	4	H	N
Polygalaceae	*Badiera penaea*	3	S / T	N
Polygalaceae	*Phlebotaenia cowellii*	4	T	E
Polygalaceae	*Polygala hecatantha*	5	H	N
Polygonaceae	*Antigonon leptopus*	1	V	I
Polygonaceae	*Coccoloba diversifolia*	25	S / T	N
Polygonaceae	*Coccoloba krugii*	15	S / T	N
Polygonaceae	*Coccoloba microstachya*	9	S / T	N
Polygonaceae	*Coccoloba swartzii*	1	T	N
Polygonaceae	*Coccoloba uvifera*	3	T	N
Polygonaceae	*Coccoloba venosa*	1	S / T	N
Portulacaceae	*Portulaca halimoides*	2	H	N
Portulacaceae	*Portulaca oleracea*	7	H	N
Portulacaceae	*Portulaca pilosa*	2	H	N
Portulacaceae	*Portulaca quadrifida*	1	H	N
Portulacaceae	*Portulaca rubricaulis*	9	H	N
Pteridaceae	*Adiantum fragile*	1	F	E
Pteridaceae	*Adiantum tenerum*	1	F	N
Pteridaceae	*Cheilanthes microphylla*	3	F	N
Rhamnaceae	*Colubrina arborescens*	7	T	N
Rhamnaceae	*Colubrina elliptica*	12	T	N
Rhamnaceae	*Colubrina verrucosa*	3	T	N
Rhamnaceae	*Gouania lupuloides*	2	V	N
Rhamnaceae	*Gouania polygama*	2	V	N
Rhamnaceae	*Krugiodendron ferreum*	13	T	N
Rhamnaceae	*Reynosia guama*	1	S / T	N
Rhamnaceae	*Reynosia uncinata*	12	S / T	N
Rhamnaceae	*Reynosia vivesiana*	5	S / T	E
Rhamnaceae	*Ziziphus mauritiana*	2	T	I
Rhamnaceae	*Ziziphus reticulata*	5	T	N
Rhizophoraceae	*Rhizophora mangle*	1	T	N
Rubiaceae	*Catesbaea melanocarpa*	1	S	N
Rubiaceae	*Catesbaea parviflora*	1	S	N
Rubiaceae	*Chiococca alba*	2	V	N
Rubiaceae	*Diodia apiculata*	3	H	N
Rubiaceae	*Erithalis fruticosa*	16	S	N
Rubiaceae	*Ernodea littoralis*	3	S	N
Rubiaceae	*Exostema caribaeum*	13	T	N
Rubiaceae	*Guettarda elliptica*	8	T	N
Rubiaceae	*Guettarda krugii*	16	T	N
Rubiaceae	*Guettarda scabra*	2	T	N
Rubiaceae	*Machaonia portoricensis*	9	S / T	E
Rubiaceae	*Margaritopsis microdon*	2	S	N
Rubiaceae	*Mitracarpus maxwelliae*	5	H	E
Rubiaceae	*Mitracarpus polycladus*	11	H	N
Rubiaceae	*Mitracarpus portoricensis*	1	S	N
Rubiaceae	*Psychotria nervosa*	2	S	N
Rubiaceae	*Psychotria pubescens*	1	S	N
Rubiaceae	*Randia aculeata*	4	T	N

Rubiaceae	*Randia portoricensis*	4	T	E
Rubiaceae	*Rondeletia inermis*	3	S	E
Rubiaceae	*Scolosanthus versicolor*	1	S	E
Rubiaceae	*Spermacoce confusa*	6	H	N
Rubiaceae	*Spermacoce remota*	1	H	N
Rubiaceae	*Stenostomum acutatum*	17	T	N
Rubiaceae	*Stenostomum lucidum*	3	T	N
Rubiaceae	*Strumpfia maritima*	10	S	N
Ruscaceae	*Sansevieria concinna*	1	H	I
Ruscaceae	*Sansevieria cylindrica*	1	H	I
Ruscaceae	*Sansevieria hyacinthoides*	3	H	I
Rutaceae	*Amyris elemifera*	18	S / T	N
Rutaceae	*Citrus* x *aurantifolia*	1	S / T	I
Rutaceae	*Citrus* x *aurantium*	1	T	I
Rutaceae	*Zanthoxylum flavum*	7	T	N
Rutaceae	*Zanthoxylum martinicense*	1	T	N
Rutaceae	*Zanthoxylum monophyllum*	4	S / T	N
Rutaceae	*Zanthoxylum spinifex*	2	S / T	N
Salicaceae	*Casearia aculeata*	1	S	N
Salicaceae	*Samyda dodecandra*	6	S	N
Salicaceae	*Xylosma buxifolia*	2	T	N
Santalaceae	*Dendrophthora brachylepsis*	1	S	N
Santalaceae	*Phoradendron anceps*	22	S	N
Santalaceae	*Phoradendron berteroanum*	1	S	N
Santalaceae	*Phoradendron quadrangulare*	1	S	N
Santalaceae	*Phoradendron* sp.	1	S	N
Santalaceae	*Phoradendron trinervium*	8	S	N
Sapindaceae	*Allophylus racemosus*	1	T	N
Sapindaceae	*Hypelate trifoliata*	8	T	N
Sapindaceae	*Melicoccus bijugatus*	2	T	I
Sapindaceae	*Serjania polyphylla*	7	V	N
Sapindaceae	*Thouinia portoricensis*	25	T	E
Sapotaceae	*Chrysophyllum oliviforme*	1	S / T	N
Sapotaceae	*Manilkara pleeana*	1	T	E
Sapotaceae	*Sideroxylon foetidissimum*	2	T	N
Sapotaceae	*Sideroxylon obovatum*	9	T	N
Sapotaceae	*Sideroxylum salicifolium*	2	T	N
Schoepfiaceae	*Schoepfia obovata*	6	S	N
Scrophulariaceae	*Capraria biflora*	2	H	N
Smilacaceae	*Smilax coriacea*	1	V	N
Solanaceae	*Datura inoxia*	1	H	N
Solanaceae	*Solanum bahamense*	4	S	N
Solanaceae	*Solanum elaeagnifolium*	2	H	N
Solanaceae	*Solanum erianthum*	1	S / T	N
Surianaceae	*Suriana maritima*	5	S	N
Talinaceae	*Talinum fruticosum*	2	H	N
Talinaceae	*Talinum paniculatum*	2	H	N
Tectariaceae	*Tectaria heracleifolia*	1	F	N
Thelypteridaceae	*Thelypteris guadalupensis*	1	F	N
Theophrastaceae	*Bonellia umbellata*	1	S	N

Theophrastaceae	*Jacquinia arborea*	11	S / T	N
Theophrastaceae	*Jacquinia berteroi*	18	S / T	N
Turneraceae	*Piriqueta racemosa*	1	H	N
Turneraceae	*Turnera diffusa*	4	S	N
Ulmaceae	*Celtis iguanaea*	2	V	N
Ulmaceae	*Celtis trinervia*	2	T	N
Ulmaceae	*Trema lamarckianum*	1	T	N
Urticaceae	*Cecropia schreberiana*	1	T	N
Verbenaceae	*Bouchea prismatica*	1	H	N
Verbenaceae	*Citharexylum spinosum*	2	S / T	N
Verbenaceae	*Duranta erecta*	2	S / T	N
Verbenaceae	*Lantana camara*	4	S	N
Verbenaceae	*Lantana exarata*	7	S	N
Verbenaceae	*Lantana involucrata*	5	S	N
Verbenaceae	*Lantana urticifolia*	5	S	N
Verbenaceae	*Stachytarpheta jamaicensis*	2	H	N
Verbenaceae	*Stachytarpheta strigosa*	3	H	N
Verbenaceae	*Tamonea boxiana*	7	H	N
Vitaceae	*Cissus obovata*	1	V	N
Vitaceae	*Cissus trifoliata*	4	V	N
Vitaceae	*Cissus verticillata*	1	V	N
Zamiaceae	*Zamia portoricensis*	4	S	E
Zygophyllaceae	*Guaiacum officinale*	11	T	N
Zygophyllaceae	*Guaiacum sanctum*	7	T	N
Zygophyllaceae	*Kallstroemia maxima*	1	H	N
Zygophyllaceae	*Kallstroemia pubescens*	2	H	N

[1] Life Form: C = Cactus, F = Fern, H = Herb, P = Palm, S = Shrub, Succ = Succulent, T = Tree, V = Vine
[2] Status: E = Endemic, I = Invasive, N = Native

Appendix III. Regression analysis for species – area bias.

Variable	*N*	*R^2*	*Adj R^2*
Rs	413	0.86	0.86

Regression Coefficients

Coef	*Est*	*S. E.*	*LL (95%)*	*UL (95%)*	*T*	*P-value*
Const	-0.11	0.06	-0.22	2.5 E -03	-1.92	0.0555
G_s	1.16	0.02	1.11	1.20	50.68	<0.0001

Analysis of Variance Table (SS type III)

S. V.	*SS*	*df*	*MS*	*F*	*P-value*
Model	1152.98	1	1152.98	2586.12	<0.0001
G_s	1152.98	1	1152.98	2586.12	<0.0001
Error	184.52	411	0.45		
Total	1337.51	412			

Appendix IV. Chi - square test for distribution of collection localities grid cells to the nearest road in the Guánica Forest plant collection.

Variable Distance (m)	*Observed*	*Expected*	*Statistic (χ^2)*	*P-value*
1 – 1000	61	69	17.4153	< 0.0001
> 1,001	15	41		
Total of grid cells counted	*76*	*110*		

Appendix V. Chi - square test for distribution of collection localities to Borinquen Camp in the Guánica Forest plant collection.

Variable Distance (m)	*Observed*	*Expected*	*Statistic (χ^2)*	*P-value*
1 – 2,000	8	12	5.2635	= 0. 0217
> 2,001	30	43		
Total of grid cells counted	*38*	*55*		

Appendix VI. Kolmogorov test with Chi – square adjustment for distribution of records through decades.

Variable	*Adjustment*	*v*	*n*	*Statistic – D*	*P-value*
Rdecade	Chi – square (1791)	179	10	1.00	<0.0001

Appendix VII. Kolmogorov test with Chi – square adjustment for distribution of records through the months.

Variable	*Adjustment*	*v*	*n*	*Statistic – D*	*P-value*
Rmonth	Chi – square (1797)	150	12	1.00	<0.0001

Appendix VIII. Chi - square test for distribution of collection localities through land cover types for the GFR plant collection.

Land cover category	*Observed number of cells with collection locality*	*Expected number of cells with collection locality*	*Statistic (χ^2)*	*P-value*
Closed forest on limestone	40	66	15.5893	= 0.0161
Open forest on limestone	15	19		
Closed forest non-calcareous	6	8		
Open forest non-calcareous	5	5		
Non forest wetland	5	7		
Grassland	8	15		
Cacti	7	6		
Total of grid cells counted	*86*	*126*		