GEOGRAPHIC ANALYSIS FOR SUPPORTING CONSERVATION STRATEGIES OF CROP WILD RELATIVES

by

NORA PATRICIA CASTAÑEDA ÁLVAREZ

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Biosciences
College of Life and Environmental Sciences
The University of Birmingham
March 2016

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Crop wild relatives are important for agriculture due to the genetic richness they possess. They have been used in plant breeding to develop high yielding varieties; varieties with improved resistance to biotic and abiotic stresses, and enhanced nutritional content. Securing their conservation in the long-term is critical to enable the continuous development of crops' varieties able to respond to future challenges. The work presented in this thesis is a contribution to the effort of understanding the *ex situ* conservation gaps of crop wild relatives, their expected response to climate change and their needs for conservation. Methods used in this thesis include species distribution modelling, gap analyses, a case study assessing the preliminary IUCN Red List categories, species distribution projections onto future climate change scenarios, and an estimation of the global value of crop wild relatives based on their likelihood of being used in plant breeding, and the contributions of their associated crops to human diets and agricultural production systems. The methods used here can be applied to more crop gene pools for global conservation planning, and can also be adapted for analysis at the regional and national level. The results presented here are being used to improve the conservation of the wild relatives of 29 crops.

To Nora, Elías and Fernando

ACKNOWLEDGEMENTS

First, I would like to thanks Dr. Andy Jarvis, who gave me the opportunity of being part of his team, and who has constantly challenged me to prove I can do better.

Special thanks to my supervisor, Dr. Nigel Maxted for his support, guidance and patience during this PhD.

My gratitude goes to Luigi Guarino, Hannes Dempewolf and Jane Toll from the Global Crop Diversity Trust, for the financial support, but also for being an inspiration to continue working towards the conservation of plant genetic resources. I also extend my thanks to Ruth Eastwood and Jonas Müller from the Millennium Seed Bank, Kew, for their support during the completion of this study.

Special thanks to the International Center for Tropical Agriculture (CIAT) and the team that has been always willing and available to help and collaborate: Chrystian Camilo Sosa, Harold A. Achicanoy, Steven Sotelo, Edward Guevara, Shirley Calderón, Ingrid Vanegas, Vivian Bernau, Ovidio Rivera, David Arango, Hugo Dorado and Carlos Navarro-Racines. Special thanks to Colin who has been my “partner in crime” during the last four years.

My gratitude also goes to all the fantastic and inspiring people I met during the preparation of this thesis: David Spooner (University of Wisconsin), Alberto Salas, Stef de Haan, Henry Juárez, Bettina Heider and Reinhard Simon (International Potato Center); Sandy Knapp, Tiina Särkinen and Mindy Syfert from the National History Museum, London; and all the genebank and database managers and herbaria curators that facilitated access to the data they maintain, specially those from the herbaria I visited personally: CUVC (Universidad del Valle, Cali, Colombia); JABOT and GUA (Rio de Janeiro, Brazil); MA (Madrid, Spain); LISC, LISI and LISU (Lisboa, Portugal); COI (Coimbra, Portugal), and E (Edinburgh, UK).

Special thanks to Dr. Daniel Debouck and Dr. Mauricio Parra-Quijano. Both of them have guided me one a way or another before starting and during this PhD.

Thanks to my friends, who have been supportive during the process of completing this PhD: Carolina González, Julie Hernández, Sergio Angulo and Carolina Navarrete. Special thanks to Meike Andersson and Julian Ramirez-Villegas for their constant support, and for putting aside some of their free time to proof-read parts of this thesis.

Profound thanks to Sandy Knapp for letting me be part of her team during the time spent at the National History Museum: these periods were an inspiration to continue working towards the understanding of plants and their conservation needs.

Thanks to the Parker family for having me during my visits to London, to Richard Barrie for the cuppas and for the Sundays when we kneaded bread, and to Isabella Römer, Paulo Ávila, Carlos Flores, Javier Juárez, and Aremi Contreras for their companion and friendship.

I am specially thankful to Marcela Quintero for all the support, counselling and coaching provided during the final stages of this PhD. Special thanks to Paul Struik who has been always available to provide insightful comments of my writing and the way I present information.

I also extend my gratitude to Sara Oldfield OBE and Dr. Eugenio Sanchez-Moran for accepting being my reviewers. I did enjoy our discussions during my viva.

This work was undertaken as part of the initiative "Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives" which is supported by the Government of Norway. The project is managed by the Global Crop Diversity Trust with the Millennium Seed Bank of the Royal Botanic Gardens, Kew UK and implemented in partnership with national and international genebanks and plant breeding institutes around the world. For further information, go to the project website: `http://www.cwrdiversity.org/`

CONTENTS

LIST OF FIGURES

LIST OF TABLES

SUPPLEMENTARY DATA

Supplementary data is available on the CD included at the back of this thesis.

Supplementary Table 2.1.

List of authors in the Crop Wild Relative Occurrence Data Global Consortium.

Supplementary Table 2.2.

List of sources of crop wild relative's occurrence data.

Supplementary Table 2.3.

List of extended ecogeographic data descriptors. Source: Castañeda-Álvarez *et al.* (2011).

Supplementary Table 2.4.

List of the data descriptors in the database.

Supplementary Table 3.1.

List of 172 species following CIP taxonomy, its equivalences in Solanaceae Source Taxonomy (Spooner *et al.*, 2014) and the prioritization category obtained through the gap analysis. SRS: Sampling Representativeness Score, GRS: Geographical Representativeness Score, ERS: Environmental Representativeness Score, FPCAT: Final priority category.

Supplementary Table 3.2.

List of bioclimatic variables (Nix, 1986) used as environmental drivers to produce environmental niche models. C.V.: coefficient of variation.

Supplementary Table 3.3.

High priority species for further collecting and the main factors contributing to insufficient representation in germplasm collections.

Supplementary Table 3.4.

List of regions and localities where further collecting may be targeted per species.

Supplementary Figure 3.1.

Boxplots showing the values obtained for the Gap Analysis metrics. Sampling Representativeness Score (SRS), Geographic Representativeness Score (GRS) and Ecosystem Representativeness Score (ERS), ordered by high priority species (HPS), medium priority species (MPS), low priority species (LPS), and "no further collecting required" (NFCR).

Supplementary Figure 3.2.

Share of species per prioritization category by taxonomic classification system. High priority species (HPS), medium priority species (LPS), low priority species (LPS), and "no further collecting required" (NFCR).

Supplementary File 3.1.

Species richness map for further exploration in Google Earth.

Supplementary File 3.2.

Potential hotspots for further collecting of high priority species (HPS) for further exploration in Google Earth.

Supplementary Table 4.1.

List of occurrence data providers used for the eggplant study.

Supplementary Table 4.2.

Adapted summary of Criterion B used to evaluate threatened categories in the form of either EOO and/or AOO (IUCN, 2012).

Supplementary Table 4.3.

Conservation status of all eggplant wild relatives used in the study. Species are list in alphabetical order. Extent of Occurrence (EOO) and Area of Occupancy (AOO) calculations described in the text. Criteria follow IUCN (2012) and Supplementary Table 4.2. Taxa assessed as threatened or near-threatened are in bold face type.

Supplementary Figure 4.1.

Scatter plots displaying the gap analysis metrics assessed for eggplant CWR. a) Sampling Representativeness Score (SRS); b) Geographic Representativeness Score (GRS); c) Ecological Representativeness Score (GRS). Black dotted lines represent the one-to-one line, which is the ideal representativeness in germplasm collections. Blue dotted lines represent a linear regression of the mean representativeness across all assessed CWR.

Supplementary Figure 4.2.

Further collecting priorities for: a) Eggplant clade (10 species); b) Climbing clade (3 species); c) Anguivi grade (36 species); d) New World relatives (3 species).

Supplementary Table 5.1

List of crops analyzed. FPS = Final Priority Score for further collecting of crop wild relatives, representing the mean FPS (± SD) across wild relatives associated with each crop. The crop importance score displays the significance of crops averaged across four global aggregate food supplies and three agricultural production metrics (see Supplementary Methods), on a scale of zero to ten, with ten representing the most important crop. These metrics are also provided in the table. ITPGRFA MLS = crop included in the Multilateral System (Annex I) of the International Treaty for Plant Genetic Resources for Food and Agriculture (FAO, 2009). Global CWR Project = target crop gene pool for activities taking place under the "Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives" initiative (Dempewolf *et al.*, 2013), ‡ Only one wild relative taxon was analyzed for the crop, § All wild relative taxa associated with this crop were rated with the same priority score.

Supplementary Table 5.2

Gap analysis results for crop wild relatives. Potential distribution models assessment scores: ATAUC = five-fold average area under the ROC curve of test data, STAUC = standard deviation of the test AUC of the five different folds, ASD15 = proportion of the potential distribution model ensemble with standard deviation above 0.15. SRS = sampling representativeness score, GRS = geographic representativeness score, and ERS = ecological representativeness score (ERS). The Final Priority Score (FPS) is the mean of SRS, GRS, and ERS. Note taxa that are members of more than one crop gene pool are listed separately in all associated gene pools, and gap analysis metrics may vary slightly for these taxa across gene pools as they were assessed separately in each case.

Supplementary Table 5.3

List of providers of occurrence records used in potential distribution modelling and gap analyses.

Supplementary Table 5.4

List of experts that evaluated the gap analysis results.

Supplementary Table 5.5

List of bioclimatic variables used as inputs in potential distribution modelling of crop wild relative taxa.

Supplementary Figure 5.1

Gap analysis metrics. a) Sampling Representativeness Score (SRS), b) Geographic Representativeness Score (GRS), and c) Ecological Representativeness Score (ERS). Gray dots represent the score obtained for each taxon. The blue dashed line represents the ideal scenario of comprehensive representation in genebanks, while the red dashed line displays the average trend across wild relative taxa.

Supplementary Figure 5.2

Collecting priorities for crop wild relatives and importance of associated crops by crop type. The priority scale displays the average of Final Priority Scores (FPS) for further collecting across wild relatives per crop. The mean importance class of associated crops displays the significance of crops averaged across four global aggregate food supplies and three agricultural production metrics (see Supplementary Methods). For both axes, the scale is zero to ten, with ten representing the highest priority for further collecting/most important crop. The size of crop gene pool circles denotes the number of wild relative taxa per crop, ranging from 1 (faba bean) to 135 (cassava).

Supplementary Figure 5.3

Gap analysis results and expert evaluation scores for prioritizing wild relatives for further collecting. a) Agreement between further collecting prioritization assigned by experts based solely upon their knowledge of gaps in genebank collections [comparable expert priority score (EPS)] and the gap analysis final priority score (FPS), assessed independently and shown as an average across wild relatives per crop gene pool. b) Agreement between further collecting prioritization degree assigned by experts based on their full knowledge of wild relatives (including threats to taxa *in situ* as well as relative value of wild relatives in crop breeding) (contextual EPS) and the gap analysis FPS. c) Qualitative expert agreement with gap analysis FPS as an average across wild relatives per crop gene pool.

Supplementary File 5.1

Supplementary methods.

Supplementary Table 6.1.

List of crop wild relative taxa and the estimated impacts of climate change on their distributions. Environmental niche models were produced for taxa with more than ten georeferenced records. Models with AUC > 0.7 were considered to assess the impact of climate change on the distributions of crop wild relatives.

LIST OF PUBLICATIONS AND PRESENTATIONS

During the preparation of this thesis, I also contributed to other published studies, different to that presented in this thesis. My contributions to these studies took place during the course of this PhD.:

Peer reviewed articles published:

Kantar MB, Sosa CC, Khoury CK, **Castañeda-Álvarez NP**, Achicanoy HA, Bernau V, Kane N, Marek L, Sieler G, and Rieseberg LH (2015). Ecogeography and utility to plant breeding of the crop wild relatives of sunflower (*Helianthus annuus* L.). Frontiers in Plant Science. doi: `10.3389/fpls.2015.00841`.

Khoury CK, Heider B, **Castañeda-Álvarez NP**, Achicanoy H, Sosa CC, Miller RE, Scotland RW, Wood JR, Rossel G, Eserman LA, Jarret RL, Yencho GC, Bernau V, JuÃąrez H, Sotelo S, de Haan S, Struik PC (2015). Distributions, *ex situ* conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., I. series Batatas]. Frontiers in Plant Science. 6:251. doi: `10.3389/fpls.2015.00251/abstract`.

Khoury CK, **Castañeda-Álvarez NP**, Achicanoy H, Sosa CC, Bernau V, Kassa MT, Norton SL, van der Maesen LJG, Upadhyaya HD, Ramirez-Villegas J, Jarvis A, and Struik PC (2015). Crop wild relatives of pigeonpea [*Cajanus cajan* (L.) Millsp.]: distributions, *ex situ* conservation status, and potential genetic resources for abiotic stress tolerance. Biological Conservation. pp. 259-270. doi: `10.1016/j.biocon.2015.01.032`.

Cobben MMP, van Treuren R, **Castañeda-Álvarez NP**, Khoury CK, Kik C, and van Hintum TJL (2015). Robustness and accuracy of Maxent niche modelling for *Lactuca* species distributions in light of collecting expeditions. Plant Genetic Resources: conservation and utilization. `doi:10.1017/S1479262114000847`.

Vincent H, Wiersema J, Kell S, Fielder H, Dobbie S, **Castañeda-Álvarez NP**, Guarino L, Eastwood R, León B, and Maxted N (2013). A prioritized crop wild relative inventory to help underpin global food security. Biological Conservation, 167, 265-275. doi: `10.1016/j.biocon.2013.08.011`.

Book chapters:

Castañeda-Álvarez NP, Khoury CK, Sosa, CC, Achicanoy HA, Bernau V, Vincent H, Jarvis A, Struik PC, Maxted N. (2016) Chapter 13: The distributions and *ex situ* conservation of crop wild relatives: a global approach. In N Maxted, ME Dulloo, and BV Ford-Lloyd (Eds.), Enhancing Crop Genepool Use: Capturing Wild Relative and Landrace Diversity for Crop Improvement. CABI, Wallingford (UK) `http://www.cabi.org/bookshop/book/9781780646138`.

Vincent H, **Castañeda-Álvarez NP**, Maxted N (2016) An approach for *in situ* gap analysis and conservation planning on a global scale. In N Maxted, ME Dulloo, and BV Ford-Lloyd (Eds.), Enhancing Crop Genepool Use: Capturing Wild Relative and Landrace Diversity for Crop Improvement. CABI, Wallingford (UK) `http://www.cabi.org/bookshop/book/9781780646138`.

Maxted N, Amri A, **Castañeda-Álvarez NP**, Dias S, Dulloo ME, Fielder H, Ford-Lloyd BV, Iriondo JM, Magos Brehm J, Nilsen L-B, Thormann I, Vincent E, Kell SP (2016) Joining up the dots: a systematic perspective of crop wild relative conservation and use. In N Maxted, ME Dulloo, and BV Ford-Lloyd (Eds.), Enhancing Crop Genepool Use: Capturing Wild Relative and Landrace Diversity for Crop Improvement. CABI, Wallingford (UK) `http://www.cabi.org/bookshop/book/9781780646138`.

Castañeda-Álvarez NP, Vincent H, Kell SP, Eastwood RJ, Maxted, N. (2011). Chapter 14 : Ecogeographic surveys. In L. Guarino, R. V Ramanatha, and E. Goldberg (Eds.), Collecting Plant Genetic Diversity: Technical Guidelines - 2011 Update. Rome, Italy. Retrieved from: `http://cropgenebank.sgrp.cgiar.org/index.php?option=com_content&view=article&id=390&Itemid=557`.

Maxted N, **Castañeda-Álvarez NP**, Vincent H, Magos Brehm J (2011). Chapter 41 : Gap analysis : A tool for genetic conservation. In L. Guarino, R. V Ramanatha, and E. Goldberg (Eds.), Collecting Plant Genetic Diversity: Technical Guidelines - 2011 Update. Rome, Italy. Retrieved from: `http://cropgenebank.sgrp.cgiar.org/index.php?option=com_content&view=article&id=390&Itemid=557`.

Presentations at conference and meetings:

Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, Bernau V, Dempewolf H, Eastwood RJ, Guarino L, Harker RH, Jarvis A, Maxted N, Müller JV, Ramirez-Villegas JA, Sosa CC, Struik PC, Vincent H, and Toll J (2015). "Conservation priorities of crop wild relative". Crop Wild Relatives Project Partners Meeting, 21st October 2015, Izmir, Turkey.

Castañeda-Álvarez NP, Khoury CK, Sosa CC, Achicanoy HA, Bernau V, Vincent H, Jarvis A, Struik PC and Maxted N (2015) "Los parientes silvestres de cultivos: Qué falta por conservar?". Corpoica, Seminario Interno, 5 August 2015, Palmira Colombia.

Castañeda-Álvarez NP (2015) "Global Conservation Priorities for Crop Wild Relatives" International Center for Tropical Agriculture (CIAT) Board Meeting, 19 May 2015, Palmira, Colombia.

Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, Sosa CC, Ramírez-Villegas J., Guarino L., Jarvis A., Maxted N (2014). "Climate change impact assessment on the distributions of crop wild relatives: a global perspective" American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America Annual International Meetings, 2-5 November 2014, Long Beach, USA. Available online at: `https://scisoc.confex.com/scisoc/2014am/webprogram/Paper88052.html` (oral presentation)

Castañeda-Álvarez NP, Khoury CK, Sosa CC, Eastwood RJ, Harker R, Vincent H, Achicanoy HA, Bernau V, Maxted N, Jarvis A (2014). "A global database for the distributions of crop wild relatives". Poster prepared for The Science Symposium of the 21st GBIF Governing Board (GB21) Meeting. New Delhi, India. 16-18 September 2014. Available online at: `http://dx.doi.org/10.6084/m9.figshare.1187065` (poster)

Castañeda-Álvarez NP, Khoury CK, Sosa CC, Achicanoy HA, Bernau V, Vincent H, Jarvis A, Struik PC, Maxted N (2014). "The distributions and *ex situ* conservation concerns of crop wild relatives: a global perspective". National History Museum Life Science Seminar. 25 June 2014. London, UK. (oral presentation)

Castañeda-Álvarez NP, Khoury CK, Sosa CC, Achicanoy HA, Bernau V, Vincent H, Jarvis A, Struik PC, Maxted N (2014). "The distributions and *ex situ* conservation concerns of crop wild relatives: a global approach". Presentation to PGR SECURE conference, 16-20 June 2014, Cambridge, UK. (oral presentation)

Castañeda-Álvarez NP (2014). "Biodiversidad como estrategia para la adaptación al cambio climático". Presentation to Universidad Nacional de Colombia, Sede Palmira, 4 April 2014, Palmira, Colombia. (oral presentation)

Castañeda-Álvarez NP, Khoury CK, Sosa CC, Achicanoy HA, Bernau V, Vincent H, Jarvis A, Struik PC, Maxted, N. (2014). "Distributions and *ex situ* conservation concerns of the crop wild relatives occurring in Brazil". Presentation to CENARGEN, April 2014, Brasilia, Brazil

Castañeda-Álvarez NP, Khoury CK, Sosa CC, Achicanoy H, Bernau V, Vincent H, Jarvis A, Struik PC, and Maxted N (2013). "A Global Perspective on Crop Wild Relatives: Distributions and Conservation *Ex Situ*" American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America Annual International Meetings, 3-6 November 2013, Tampa, USA. Available online at: `https://dl.sciencesocieties.org/publications/meetings/2013am/11558` (oral presentation)

Castañeda N (2013). "Gap analysis of crop wild relatives" and "Accessing and downloading biological records from public databases: GBIF and GENESYS". Capacity building workshops for the strengthening of the capacities of the national programs on plant genetic resources in Latin America (CAPFITOGEN), March 2013, Bogotá, Colombia. (oral presentation)

DECLARATION

The work presented in Chapter 3 has been published in PLoS ONE. The work presented in Chapters 4 and 5 has been accepted for publication. The work in Chapters 2, 6 and 7 is presented in a suitable format for publication and has not been yet submitted for publication. The content of all chapters is largely identical to the manuscripts presented for publication. All chapters were written by me and represent my work, with the following exceptions: Chapter 3 was written in collaboration with Dr. Stef de Haan, Chapter 4 was largely written by Dr. Sandra Knapp (corresponding and senior author) and Dr. Mindy M. Syfert, with substantial contributions from me. Chapter 5 was written collaboratively with Dr. Colin K. Khoury. Dr. Sandra Knapp, Dr. Mindy M. Syfert, Dr. Stef de Haan and Dr. Colin K. Khoury have agreed with me to use their contributions in this thesis. All have signed the declaration presented at the end of this section. Chapter 2 had contributions from multiple authors as indicated at the start of Chapter 2. All figures and tables were prepared by me, with specific exceptions described in the contributions note included in this section.

Here a list of the chapters published, in press, submitted and to be submitted:

Chapter 2:

Castañeda-Álvarez NP, Khoury CK, Sosa CC, Sotelo S, Vanegas I, Calderón S, Vincent H, Harker RH, Bernau V, Eastwood RJ, Ramirez-Villegas J, Dempewolf H, Guarino L, Toll J, Jarvis A, Müller J, the Global Consortium of Crop Wild Relative Occurrence Data Providers, Maxted N. (to be submitted) A global occurrence dataset for crop wild relatives. *Scientific Data*.

See contributions note below

Chapter 3:

Castañeda-Álvarez NP, de Haan S, Juárez H, Khoury CK, Achicanoy HA, Sosa CC, Bernau V, Salas A, Heider B, Simon R, Maxted N, Spooner, DM. (2015) *Ex situ* conservation priorities for the wild relatives of potato (*Solanum* L. section *Petota*). *PLoS ONE* 10(4): e0122599. **See contributions note below**

Chapter 4:

Syfert MM, Castañeda-Álvarez NP, Khoury CK, Särkinen T, Sosa CC, Achicanoy HA, Bernau V, Prohens J, Daunay M-C, Knapp S. (in press) Crop wild relatives of the brinjal eggplant (*Solanum melongena*: Solanaceae): poorly represented in genebanks and many species at risk of extinction. *American Journal of Botany*. **See contributions note below**

Chapter 5:

Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, Bernau V, Dempewolf H, Eastwood RJ, Guarino L, Harker RH, Jarvis A, Maxted N, Müller JV, Ramirez-Villegas J, Sosa CC, Struik PC, Vincent H, Toll J. (in press) Global conservation priorities for crop wild relatives. *Nature Plants*. **See contributions note below**

Contributions to Chapter 2

Chapter 2 had contributions from multiple authors as described at the start of the Chapter. My contributions included:

- Conceiving the study, searching and coordinating the collection of occurrence data.
- Collecting occurrence data.
- Designing the database that stored the data.
- Led the writing of the manuscript for publication.

Contributions to Chapter 3

My contributions to the text in Chapter 3 are described here:

- Introduction: Paragraphs 1, 2 (first and fifth sentences), 4, 5 (second sentence), 6, 7.
- Methods: Paragraphs 1, 2, 3, 4, 5, 6, 7, 8.
- Results: Paragraphs 1, 2, 3, 4, 6 (first, second, third, fourth sentences), 7.
- Discussion: Paragraphs 1 (first sentence), 2, 4, 5 (first, second, third and fourth sentences), 6 (first, second, fifth and sixth sentence), 7.

All other parts of the text in Chapter 3 were contributed by Dr. Stef de Haan. Table 3.1 was prepared with inputs from Dr. Stef de Haan. Figure 3.1 was prepared by Dr. Stef de Haan.

Contributions to Chapter 4

I contributed and actively participated in the conception and design of the work presented in Chapter 4. This includes the collation and preparation of data, the *ex situ* gap analysis, and the writing of the text. My contributions to the text are reflected in these sections:

- Introduction: Paragraphs 2 (second sentence), 3, 4 (second and third sentence), 5 (first sentence), 6 (fourth and fifth sentence).
- Methods: Paragraphs 1(first sentence), 2 (fifth sentence), 5, 6 (fourth sentence), 7.
- Results: Paragraphs 2 (first sentence), 5 (first sentence), 6 (first sentence), 8 (first sentence).
- Discussion: Paragraphs 1(fifth sentence), 5 (first and third sentence), 11 (first, second and fifth sentence).

All other parts of the text in Chapter 4 were contributed by Dr. Sandra Knapp and Dr. Mindy M. Syfert. Here a description of my contributions in relation to the figures and tables of this chapter:

- Table 4.1: I prepared most of this table and Dr. Jaime Prohens (co-author) gave some inputs.
- Table 4.2: This table was prepared collaboratively with Dr. Mindy M. Syfert and Dr. Sandra Knapp.
- Table 4.3: Prepared by Dr. Mindy M. Syfert.
- Figure 4.1 and Figure 4.3: Dr. Mindy M. Syfert prepared the figures, I contributed to collect and organize the data presented in these figures.
- Figure 4.2: Dr. Mindy M. Syfert prepared the figure, I contributed in processing the data displayed in the figure.

- Figure 4.4: Dr. Mindy M. Syfert prepared the figure.

Contributions to Chapter 5

Chapter 5 was written collaboratively with Dr. Colin K. Khoury, with whom I share the first authorship of this paper. Here the list of my contributions to the text:

- Introduction: Paragraph 1 (second and fourth sentence), 2 (first, second and fourth sentence), 3.
- Methods (Supplementary methods): Paragraph 1, 2 (first, third and fourth sentence), 3 - 7, 8 (first, second, third and sixth sentence), 9 (first and second sentence), 10, 11.
- Results: Paragraph 1 (second and fourth sentence), 2 - 5.
- Discussion: Paragraph 1 (first, fourth and fifth sentence), 2 (first sentence), 3 (first, second, third, fifth, sixth sentence), 4, 5 (first and third sentence).

CHAPTER 1

INTRODUCTION

1.1 Context

Agriculture is an important economic activity, contributing up to a third of the Gross Domestic Product (GDP) in low income countries, providing employment to one of three people living in low and lower income countries and is a driving force of rural development in least-developed countries (FAO *et al.*, 2012; The World Bank, 2014a,b). Moreover, agriculture is the basis of food security and nutrition.

Multiple studies had shown that agriculture will be negatively affected by climate change (Challinor *et al.*, 2014; Porter *et al.*, 2014). This, in combination with a growing population demanding food, less natural resources available for food production, and the need of reducing the negative impacts of agriculture to the environment, pose the challenge of producing more food in a more sustainable manner (Foley *et al.*, 2011; Tilman *et al.*, 2011).

Having access to the total wealth of genetic diversity and using it to develop crop varieties, is one of the strategies we can rely on for making agriculture more sustainable and at the same time resilient to climate change (Foley *et al.*, 2011; Guarino and Lobell, 2011; McCouch *et al.*, 2013). Many modern cultivars have been described as having a narrow genetic base (Kannenberg and Falk, 1995; van de Wouw *et al.*, 2009), therefore securing novel sources of genes that can help expanding the genetic base of crops is warranted. Crop wild relatives (CWR) have not undergone the genetic bottleneck as domesticated species had, allowing them to carry a wide genetic diversity, and thus, they can be used as a source of genes for cultivated species (Doebley *et al.*, 2006; McCouch *et al.*, 2007; van de Wouw *et al.*, 2009; Fuller *et al.*, 2014). Despite their importance, crop wild relatives are currently under-represented in *ex situ* collections and their habitats are exposed to threats that can jeopardize their availability in the future (FAO, 2010; Ureta *et al.*, 2012). Understanding the current status of *ex situ* conservation, the geographical distribution patterns and the threats potentially affecting crop wild relatives can help to guide future conservation efforts, and therefore, enhance their availability for being used in crop improvement.

1.2 Relevance of agriculture in the world

Agriculture plays a key role in income generation and human nutrition globally. In 2012, Agriculture contributed 12.5% of the global Gross Domestic Product (GDP), representing up to 33% and 17% of the GDP in low and lower middle income countries respectively (The World Bank, 2014a,b). This sector provides employment to 17-30 % of the workforce in lower and upper middle income countries (see Figures 1.1 and 1.2). In addition, about 2.5 billion people (35% of world population) depend directly from agriculture for their livelihoods (IFAD and UNEP, 2013).

Agricultural systems vary around the world, from industrialized holdings growing crops for food, animal feed and biofuels, to small sized farms providing around 35% of the global production of maize, soybean, wheat and rice altogether (Syngenta, 2013). Moreover, a large share of smallholders in the world are involved in agriculture (about 85% of the farms worldwide have an extension of less than two hectares. See Figure 1.3). Smallholder agriculture plays an important role in food security and nutrition. For instance, Asia and Sub-Saharan Africa, consume 80% of the agricultural output produced by smallholders, and Latin America and the Caribbean consume 73% of the regional production (Berdegué and Fuentealba, 2011; IFAD and UNEP, 2013).

Likewise, strong linkages between agriculture and poverty alleviation have been found, with

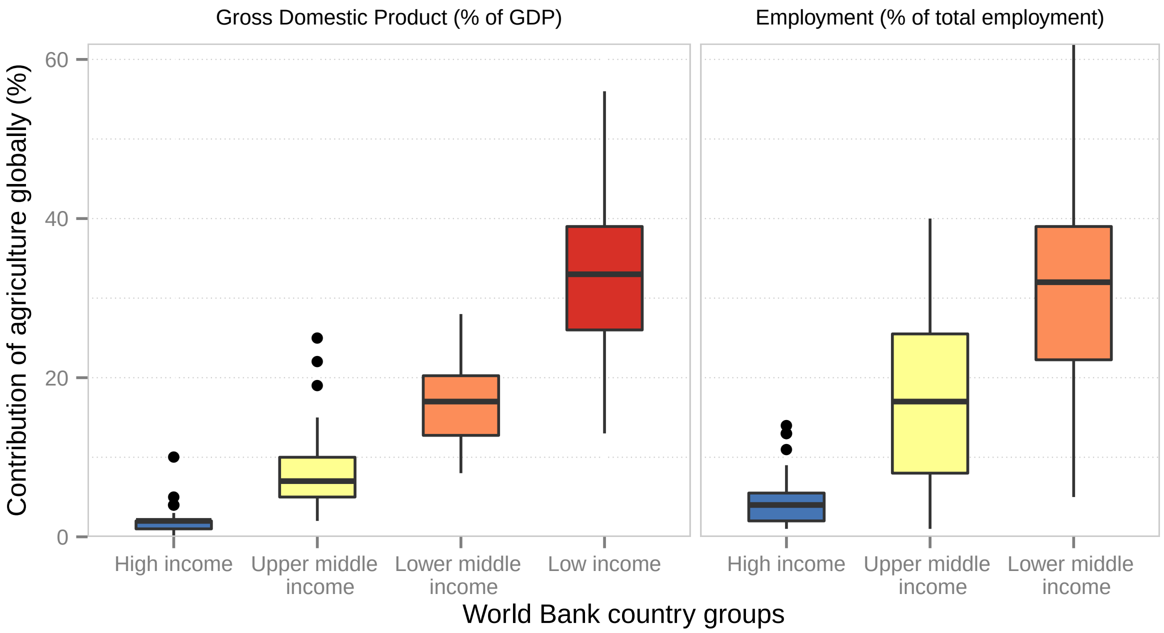


Figure 1.1: Share of agriculture in global GDP and employment, per country income group. Graphic by the author. Data source: (The World Bank, 2014a,b).

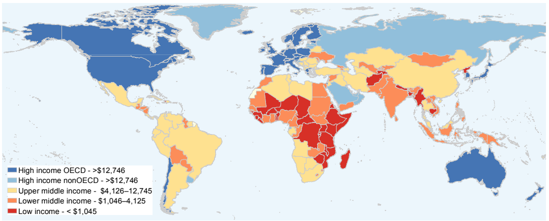


Figure 1.2: The World Bank country income group classification.

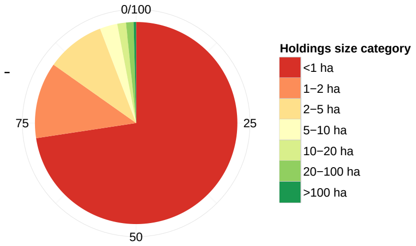


Figure 1.3: Share of farm units per size category. Graphic by the author. Data source: HLPE (2013).

noticeable differences between regions (Irz *et al.*, 2001), where least-developed countries more benefited from agriculture to reduce poverty (FAO *et al.*, 2012). Impact assessments of investments in research and development (R&D) for agricultural productivity growth have shown positive effects reducing poverty in Africa, Asia and a handful of countries in Latin America (Thirtle *et al.*, 2003). Due to the aforementioned considerations, agriculture, in particular that involving smallholders, is being promoted as a means to guarantee food security, improve nutrition and foster economic growth in the world (HLPE, 2013).

1.3 Constraints and challenges for agriculture

According to the Intergovernmental Panel on Climate Change (IPCC), Fifth Assessment Report (AR5), agriculture is considered one of the major economic activities contributing to greenhouse gas emissions (GHG), particularly non-CO_2 emissions as methane and nitrous oxide, accounting for 10-12% of the global GHG derived from human activities (Ahammad *et al.*, 2014). Main sources of such emissions are the use of manure and synthetic fertilizers and paddy rice cultivation (Figure 1.4).

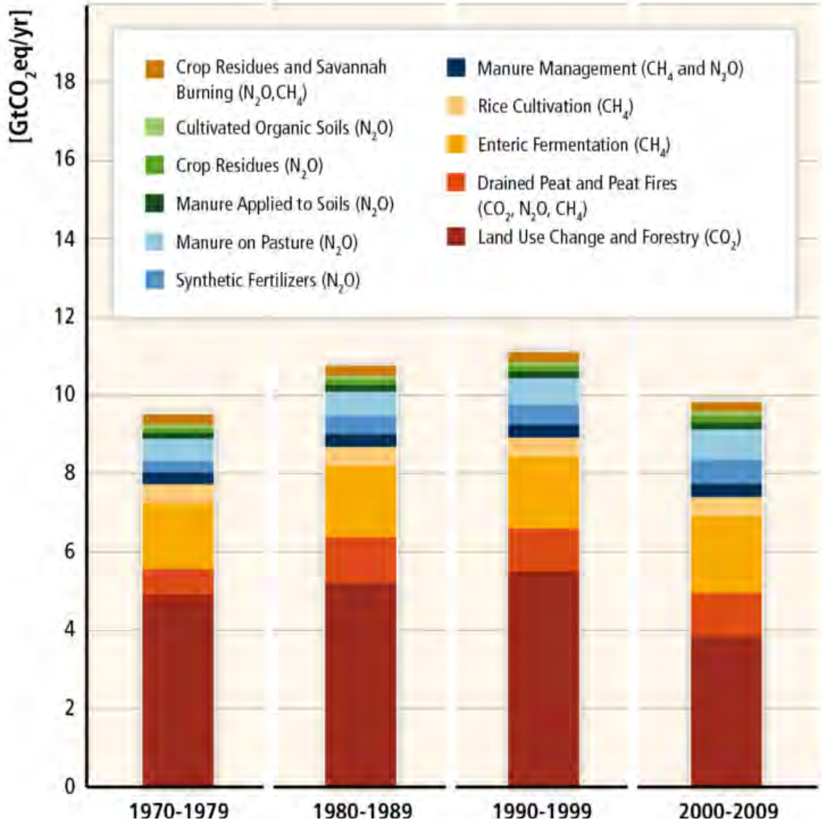


Figure 1.4: Agriculture GHG emissions in the last four decades. Source: Ahammad *et al.* (2014).

Simultaneously, agriculture is being affected by climate change. Effects of climate change like warmer temperatures, changes in precipitation patterns and an elevated concentration of CO_2 and O_3 in the atmosphere, affect negatively the productivity of crops (Ainsworth *et al.*, 2012; Ruiz-Vera *et al.*, 2013), even under the assumption that adaptation measures are implemented on time (Porter *et al.*, 2014). It is expected that by 2030, reductions of crop yields will become more evident as a consequence of higher temperatures (Porter *et al.*, 2014). In a meta-analysis, Challinor *et al.* (2014) described the impact of climate change in three major cereals (maize, wheat and rice), in temperate and tropical areas, finding that general losses of crop yields may occur with a rise of 2 °C in temperatures, and crop productivity in the tropics is expected to be comparatively more affected than crop productivity in temperate regions. Similarly, studies assessing the vulnerability of crops of importance for food security found a significant reduction in crop productivity by mid-century, as it is the case for sorghum (-17% of its current production), millet (-17%), groundnut (-18%) and cassava (-8%) (Schlenker and Lobell, 2010). Multiple studies have analyzed the impact of climate change in areas considered climatically suitable for growing bananas, beans, potatoes and cassava globally, finding an overall reduction of suitability in most cases, with the exception of cassava, where new areas meeting the climatic requirements of the crop are likely to appear towards the outer limits of the tropical belt (Beebe *et al.*, 2011; Ramirez *et al.*, 2011; Schafleitner *et al.*, 2011; Jarvis *et al.*, 2012). Besides the negative effects in crop yields and shifts in climatic suitable geographies, climate change is expected to affect the nutritional content of crops and forages (Taub *et al.*, 2008; Perring *et al.*, 2010; Myers *et al.*, 2014).

At the same time, a higher pressure on food systems is expected as human population is likely to increase to 9.6 billion people by mid-century (United Nations, 2013), along with the increasing intake of calories obtained from fewer crop commodities as wheat, rice and maize (FAO *et al.*, 2012; Khoury *et al.*, 2014), and growing consumption of animal derived products like meat, eggs and dairy as a consequence of higher incomes and purchasing power around the world (Delgado, 2003; Cordell *et al.*, 2009; Msangi and Rosegrant, 2011). Moreover, inputs such as water for irrigation, fertilizers and arable land, required for food production are becoming scarcer and/or more unevenly distributed around the world (Rosegrant *et al.*, 2002; Van Vuuren *et al.*, 2010), and the trade-offs associated with the use and exploitation of such resources, are compelling reasons to explore an enhanced and more efficient use of the resources required for crop production (Foley *et al.*, 2005; Bodirsky *et al.*, 2014).

1.4 Plant genetic resources and agriculture

Plant genetic resources (PGRs), as functional elements of biodiversity, are sources of diversity of genetic material within plants important for agriculture including plant landraces, genetic stock, primitive forms of cultivated species, modern and obsolete cultivars, breeding lines, weeds, wild relatives and unrelated species (Hawkes *et al.*, 2000). Historically, humanity has benefited from PGRs in multiple ways. Landraces of maize (*Zea mays* L.) are reported to be used to prepare traditional dishes in North and Central America (CONABIO, 2012), traditional cultivars of achiote (*Bixa orellana* L.) have been used as a natural dye since pre-Columbian times (Smith and Schultes, 1990) and *Oryza glaberrima* Steud. has been reported to be used in rituals in West Africa (Linares, 2002), just to mention few cases. Likewise, plant genetic resources are key elements of livelihoods and a safeguard strategy in case of harvest failures for some communities (Olson *et al.*, 2012; Vasconcelos *et al.*, 2013; Zimmerer, 2013)

Similarly, modern plant breeding has relied on plant genetic resources and associated allelic diversity to produce hybrids and varieties with certain characteristics as higher yields, better resistance and tolerance to pest and diseases, enhanced nutritional content, improved end-use quality and more recently, to broaden the genetic base of some crops (IRRI, 1990; Hoisington *et al.*, 1999; Gepts, 2006). Likewise, plant breeding has been considered as one of the strategies with potential to improve crop productivity, to enhance the efficient use of agricultural inputs and to adapt crops to the climatic conditions expected under climate change (Singh *et al.*, 2009; Subbarao *et al.*, 2009; Ceccarelli *et al.*, 2010; Godfray *et al.*, 2010; Foley *et al.*, 2011).

Given the importance and potential of plant genetic resources for global agriculture, different initiatives and strategies have been put in place to facilitate the access, conservation and use of such resources globally. Some remarkable milestones in this effort are: the FAO/IBP Technical Conference on the Exploration, Utilization and Conservation of Plant Genetic Resources (1976) where concerns were raised over the rapid loss of genetic diversity and discussions on long term conservation strategies for breeding (i.e., *ex situ* and *in situ* approaches) took place (Pistorius, 1997); the establishment of a global network of genebanks and further collection of 120,000 new accessions coordinated by the International Board for Plant Genetic Resources

(IBPGR) (1974-1984) (Pistorius, 1997); and the negotiations and following adoption of the International Treaty on Plant Genetic Resources for Food in Agriculture (ITPGRFA) which provides the basis for facilitating access and benefit sharing derived from the use of PGRs (FAO, 2009).

There are about 7.4 million accessions conserved *ex situ*, with a large proportion of them believed to be duplicates (65-70% of the total accessions) (FAO, 2010). However, some categories, in particular the crop wild relatives (CWR) are still inadequately represented (10% of the total holdings), especially in crop groups like food legumes (4%), fibre crops (4%), cereals (5%) and vegetables (5%) (FAO, 2010).

1.5 Crop wild relatives

1.5.1 Definition of CWR

Crop wild relatives are the wild and weedy taxa genetically related to crops, including their ancestors (Maxted *et al.*, 2006; Heywood *et al.*, 2007). Unlike their associated crops, CWR have not undergone the genetic bottleneck of domestication, dispersal and modern breeding, resulting in CWR being more genetically diverse than their cultivated counterparts, as it has been evidenced for some crops like pearl millet [*Pennisetum glaucum* (L.) R. Br.] (Mariac *et al.*, 2006), wheat (*Triticum aestivum*) (Reif *et al.*, 2005), pigeonpea [*Cajanus cajan* (L.) Millsp.] (Yadav, 2012), soybean (*Glycine max*) (Lam *et al.*, 2010), and African rice (*Oryza glaberrima*) (Nabholz *et al.*, 2014). Thanks to the genetic diversity, unique traits, adaptations to particular environmental conditions and relative easiness for crossing with cultivated species, CWR are important for plant breeding and therefore agriculture (Dale, 1992; Tanksley and McCouch, 1997).

Various classification schemes have been proposed to describe the degree of relatedness of crop wild relatives and their associated crops. Harlan and de Wet (1971) proposed the Gene Pool concept, composed by the category primary gene pool (GP-1) subdivided into GP-1a for cultivated taxa, and GP-1b for spontaneous taxa, the secondary gene pool (GP-2) for the taxa that can cross with the crop but may produce sterile hybrids, and the tertiary gene pool (GP-3) for the taxa that produces sterile hybrids when crossed with the crop, and additional biotechnological tools are required for rescuing the offspring produced after the crossing. However, hybridization essays between CWR taxa and cultivated species have been performed and reported for a handful of crops, and therefore alternative working definitions have been proposed. Maxted *et al.* (2006) proposed the Taxon Group concept as a mean to predict the degree of relatedness between CWR taxa and their associated crop based on taxonomic hierarchy. The Taxon Group concept contains the levels: Taxon Group 1a (TG1a), enclosing the cultivated species; Taxon Group 1b (TG1b), including the subspecies and varieties of the cultivated species; Taxon Group 2 (TG2), enclosing the taxa of the same series or section as the crop; Taxon Group 3 (TG3), composed by the taxa in the same subgenus as the crop; Taxon Group 4 (TG4), for the taxa of within the same genus as the crop; and Taxon Group 5 (TG5) containing the taxa from the same tribe but different genus than the crop. Wiersema *et al.* (2012) proposed the genetic relative concept. This concept uses data on phylogeny, ploidy, reproductive biology, and reports of natural hybridization events to define four ranks: the primary genetic relative status (PGR), the secondary genetic relative status (SGR), the tertiary genetic relative status (TGR), and the graftstock class for perennial species that are used as grafts. Finally, the provisional gene pool concept is used when no gene pool concept has been reported in literature, but published evidence of crossings is available (Vincent *et al.*, 2013) (Figure 1.5).

1.5.2 Utilization of CWR

CWR have been used in crop breeding as novel materials to increase the genetic diversity and broaden the genetic base of crops, and as sources of various traits, some examples include: resistance to coffee leaf rust in *Coffea arabica* derived from *C. canephora* and *C. liberica* (Anthony *et al.*, 2011), resistance against rust and late leaf spot in peanut (*Arachis hypogaea*) obtained from *A. cardenassi* Krapov. & W.C. Gregory (Moss *et al.*, 1997), tolerance to corn rootworms in *Zea mays* L. through crossings with *Tripsacum dactyloides* L. (Prischmann *et al.*, 2009),

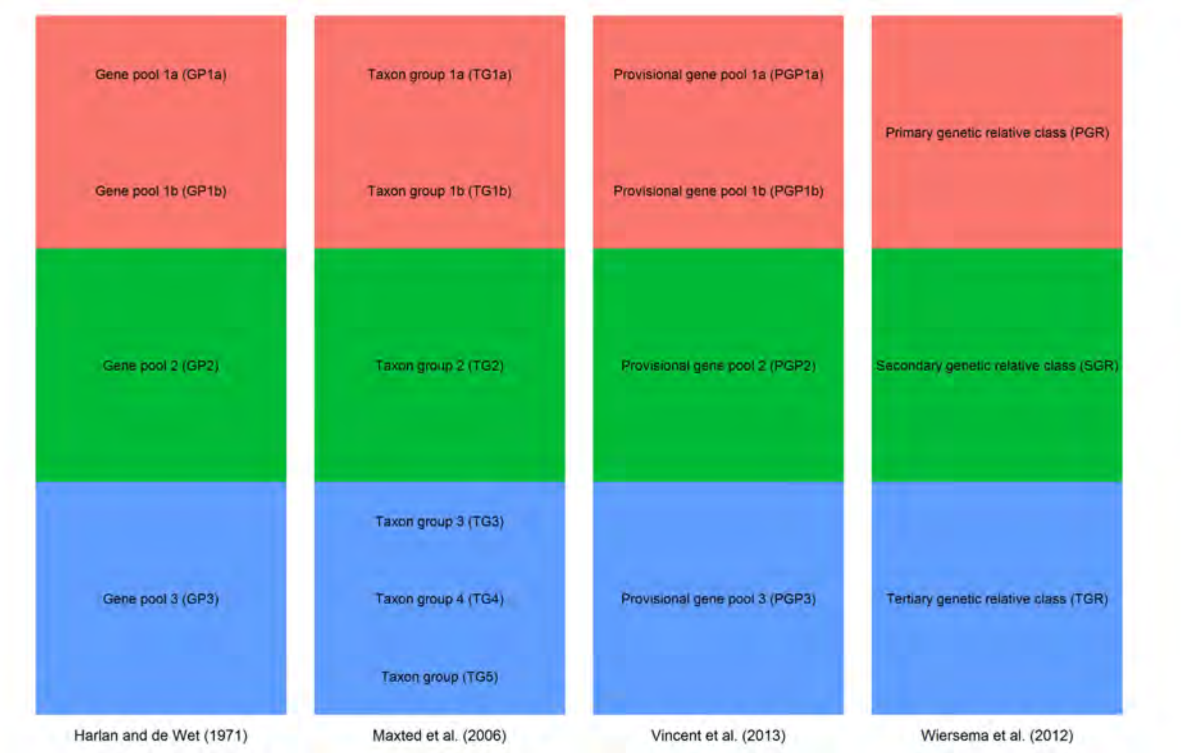


Figure 1.5: Classification schemes of the degree of relatedness of CWR to their associated crops.

changes in the size of leaves and inflorescence in pearl millet as a result of introgressing *Pennisetum squamulatum* (Dujardin and Hanna, 1989), enhanced yields in barley by using *Hordeum vulgare* subsp. *spontaneum* and chickpea using *Cicer reticulatum*, improved end-use quality of tomatoes using *Solanum chmielewskii* for increasing the soluble solids content in tomato fruits (Rick and Chetelat, 1995), improved flavors in strawberries derived from *Fragaria vesca* (Ahmadi and Bringhurst, 1992), and enhancement of the nutritional content of maize (Wusirika *et al.*, 2011), broccoli (Traka *et al.*, 2013), and cassava, beans and wheat (Pfeiffer and McClafferty, 2007).

Likewise, CWR have been used as sources of resistance to abiotic constraints like drought in wheat (Gororo *et al.*, 2002), oats (Suneson, 1967), strawberries (Ahmadi and Bringhurst, 1992), and rice (Zhang *et al.*, 2006); salinity in tomatoes (Rick and Chetelat, 1995), wheat (Farooq *et al.*, 1995) and sunflower (Miller and Seiler, 2003); and heat in rice (Ishimaru *et al.*, 2010). These particular traits make CWR perfect candidates for helping crops to adapt to the expected conditions of climatic change (Guarino and Lobell, 2011; McCouch *et al.*, 2013; Dempewolf *et al.*, 2013; Maxted *et al.*, 2014). Additionally, the cost of the technologies that can further advance the understanding of the genes controlling particular traits in CWR, the underlying processes that enable two different species to interbreed successfully, and hybridization between CWR and wild species is decreasing, and therefore it is expected that the use of CWR for producing more productive, nutritious and resilient varieties is expected to grow (Tanksley and McCouch, 1997). Furthermore, CWR pre-bred materials should progressively become more available for conventional breeding programs, playing a decisive role to continue incorporating wild genes into crops (Nass and Paterniani, 2000; Valkoun, 2001; Dempewolf *et al.*, 2013; Sharma *et al.*, 2013).

1.5.3 Threats affecting CWR

Crop wild relatives, as a group of plants, may be threatened to the same degree as other groups of living beings like mammals (Brummitt and Bachman, 2010; Brummitt *et al.*, 2015a). Using the Sampled Red List Index (SRLI), Brummitt *et al.* (2015a) found that 22% of the plant species in the world are threatened with extinction, as they are classified within the IUCN categories "Critically Endangered" (CR), "Endangered" (EN), and "Vulnerable" (VU).

The destruction, degradation and loss of habitats, species over-exploitation, pollution and invasive species, are identified as the main drivers of threats for biodiversity (Hudson *et al.*, 2014). The major threat for plants is the conversion of natural habitats to agriculture (Brummitt and Bachman, 2010). Climate change will compound the risk of extinction for many plant species, including CWR. For instance, high fragmentation and significant reductions of the ranges of CWR have been estimated for the wild relatives of potato (*Solanum*), cowpea (*Vigna*), peanut (*Arachis*) and maize (*Zea*) in a warmer world (Jarvis *et al.*, 2008; Ureta *et al.*, 2012).

The transference of genes from domesticated crops to CWR taxa has been described as an additional threat that may result in the extinction of rare CWR taxa, due to the probable loss of genetic integrity or outbreeding depression (Ellstrand *et al.*, 1999). *Oryza rufipogon* subsp. *formosana*, *Gossypium darwinii* and *G. tomentosum* have been considered under the risk of extinction due to their hybridization with their associated crops (Ellstrand *et al.*, 1999).

1.5.4 Policies supporting the conservation of CWR

Given the importance of plant genetic resources (including the wild relatives of crops), several international policy frameworks have been established to promote the conservation and use, and secure the equitable distribution derived from the use of plant genetic resources. These policy frameworks include: The Convention on Biological Diversity (CBD) (including the Nagoya Protocol), the ITPGRFA, the Second Global Plan of Action for Plant Genetic Resources for Food and Agriculture, and more recently the United Nation's Sustainable Development Goal 2, Target 2.5. Herein a brief explanation of these frameworks.

The aims of the Convention on Biological Diversity (CBD) are:

> "The conservation of biological diversity, the sustainable use of its components and the fair and equitable sharing of the benefits arising out of the utilization of genetic resources, including by appropriate access to genetic resources and by appropriate transfer of relevant technologies, taking into account all rights over those resources and to technologies, and by appropriate funding." (CBD, 1992)

The CBD is implemented through the Strategic Plan for Biodiversity 2011 - 2020, which contains five strategic goals and twenty Aichi Biodiversity Targets. The Strategic Goal C ("To improve the status of biodiversity by safeguarding ecosystems, species and genetic diversity"), comprises the Aichi Target 13, an ambitious target for conserving PGR by 2020:

> "By 2020, the genetic diversity of cultivated plants and farmed and domesticated animals and of wild relatives, including other socio-economically as well as culturally valuable species, is maintained, and strategies have been developed and implemented for minimizing genetic erosion and safeguarding their genetic diversity." (Secretariat of the CBD, 2010)

The Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity (Nagoya Protocol for short) is a supplementary agreement to the CBD to comply with the objective of sharing the benefits derived from the use of genetic resources in a fair and equitable manner (Secretariat of the CBD, 2011).

Aligned to the CBD is the ITPGRFA (or "Plant Treaty"), in force since 2004. The Plant Treaty aims to promote the conservation and use of plant genetic resources for sustainable agriculture and food security, guaranteeing at the same time access to PGR and an equitable sharing of the benefits derived from their use. For this, members of the Treaty (i.e., Contracting Parties) are called to promote the conservation, exploration, collection, characterization, evaluation and documentation of PGR, and to reduce threats that may affect them (FAO, 2009). The Treaty also devised the Multilateral System of Access and Benefit-sharing as a mechanism to facilitate the access to PGR held among the Contracting Parties and the International Network of *Ex Situ* Collections (i.e., the collections conserved in trust for the global community by the International Agricultural Research Centres of the CGIAR, and other international institutions), and securing benefit sharing, whilst reducing the transactions costs that bilateral negotiations may imply (FAO, 2009).

The Second Global Plan of Action for PGRFA (Second GPA), a supporting component of the Treaty, is a global framework for the conservation and sustainable use of PGRFA in the face of the challenges of food insecurity and climate change (CGRFA, 2012). The Second GPA is considered a reflection of the commitments of governments to promote plant genetic resources under the challenges of food insecurity and climate change and new opportunities arisen since the Global Plan of Action on Conservation and Sustainable Use of PGRFA, adopted in 1996, such as new tools available for the management of PGR, biotechnological advances for using PGR in plant breeding, and the development of products derived from agriculture (CGRFA, 2012). Its implementation is expected to greatly complement the Treaty and the achievement of the Aichi Target 13 (CGRFA, 2012).

And finally, the United Nation's Sustainable Development Goals (SDGs) include an explicit call to conserve the genetic diversity of crops in its Goal 2 ("Zero hunger"), Target 2.5:

> "By 2020, maintain the genetic diversity of seeds, cultivated plants and farmed and domesticated animals and their related wild species, including through soundly managed and diversified seed and plant banks at the national, regional and international levels, and promote access to and fair and equitable sharing of benefits arising from the utilization of genetic resources and associated traditional knowledge, as internationally agreed." (SDGs; `http://www.un.org/sustainabledevelopment/`

`hunger/`).

1.5.5 Conservation assessments for CWR

Conservation assessments are critical for defining the conservation priorities of taxa, populations and/or areas that may need to be conserved, enabling a conservation effort to use resources more efficiently, while at the same time meeting its conservation objectives.

Various methodologies have been used for assessing the conservation status of CWR, they vary according to the scope of the conservation effort and for the case of PGR, whether the planned conservation will be *in situ* or *ex situ*. Regarding the specific case of CWR, Flor *et al.* (2006) proposed a group of five criteria for prioritizing European CWR for conservation:

- Threat: Assesses the risk of extinction of taxa, or any existing and potential threat that may hamper their survival.
- Conservation: Assesses whether the taxa is already *in situ* or *ex situ* conserved, and to what extent.
- Genetic: Assesses the degree of relatedness of CWR to its associated crops, and if the taxa have been affected by genetic erosion or genetic pollution.
- Economic: Assesses the actual and potential economic importance of CWR at the local and national level.
- Utilization: Assesses the social importance of CWR taxa, the harvesting frequency, and if there is any cultural use associated to the taxa.

Ford-Lloyd *et al.* (2008) proposed a one-criteria prioritization approach for CWR, based on either one of these parameters: distribution as a proxy of threats, assuming that if a taxon is endemic to a single country it might be at higher risk of extinction, if compared to a more widespread taxon distributed in two or more countries; and a socio-economic value score based on experts' opinions. Meanwhile, Brehm *et al.* (2010) in an analysis circumscribed to Portugal used the following eight criteria for defining the conservation priorities of CWR with potential and confirmed socio-economic value: 1) Native status of the taxa (whether the taxa is considered native to the country, introduced, allien, or there is not enough information to define this category), 2) Economic value, 3) Ethnobotanic value, used to capture whether the taxa has an use associated to the local traditions and culture, 4) Distribution in the world, 5) Distribution within the country, 6) *In situ* and *ex situ* conservation status, 7) Legislation, and 8) IUCN Red Listing category.

The Gap Analysis, a widely applied methodology for identifying elements of biodiversity that are not actively conserved, has been also used to determine the conservation priorities of CWR. Maxted *et al.* (2008) designed a framework consisting in four steps for using the Gap Analysis for identifying the *in situ* and *ex situ* conservation needs of CWR:

- Step 1: Definition of target taxon and target area. Given the breadth of CWR taxa and their geographical distributions, this first step helps focusing the conservation assessment to the taxon (e.g., order, family, genus, species), and geographic scope of interest for the analysis.

- Step 2: Assessment of natural diversity. The Gap Analysis needs a baseline of the current existing diversity of the taxon to be analyzed, as it will help identifying the taxa, populations and areas unrepresented and underrepresented whether *ex situ* or *in situ*. Maxted *et al.* (2008) proposed four evaluations within this step: 1) Taxonomic diversity assessment, 2) Genetic diversity assessment, 3) Ecogeographical diversity assessment, and 4) Threat assessment.

- Step 3: Assessment of current conservation strategies. This step consists on reviewing the extent of conservation of the taxon(a) of interest within protected areas, genetic reserves and on-farm conservation initiatives for the case of *in situ* conservation. For *ex situ* conservation, an analysis of the representativeness of the taxon(a) under analysis in genebanks is required.

- Step 4: Reformulation of conservation strategy. Gaps in conservation are identified after comparing the existing diversity of the taxon(a) under study versus the extent of representativeness of such taxon(a) in *ex situ* and *in situ* conservation strategies. Once the gaps are identified, *in situ* and complementary *ex situ* conservation strategies are subsequently designed.

Maxted *et al.* (2008) applied its Gap Analysis framework to the CWR of cowpea (*Vigna unguiculata* L.) native to Africa.

A Gap Analysis methodology for determining the extent of representativeness of CWR in genebanks, and therefore defining the *ex situ* conservation priorities of CWR taxa was proposed by Ramírez-Villegas *et al.* (2010). This methodology combines some of the parameters from Maxted *et al.* (2008) with enviromental niche models (ENM) and geographical analyses, producing distribution maps of each CWR taxon under analysis, maps presenting the areas where new collections of plant material should be conducted, richness maps portraying areas where multiple CWR taxa may occur, and three quantitative metrics that describe how well represented is a CWR taxon in a genebank. These metrics are:

- Sampling Representativeness Score (SRS): This score compares the amount of existing genebank accessions of the CWR taxon under analysis with the number of reference records (e.g., herbarium specimens, inactive germplasm accessions, field observations), providing a general overview of the sufficiency of genebank accessions of the CWR taxon.

- Geographical Representativeness Score (GRS): This score assesses the current geographical adequacy of genebank accessions versus the potential environmental niche that the taxon can occupy. In order to obtain this score a circular buffer area of 50km (CA50) (Hijmans and Spooner, 2001) is calculated around each georeferenced genebank accession, and an enviromental niche model is prepared to represent the potential extent of the taxon. Environmental drivers and georeferenced occurrences are used as inputs for the modelling algorithm of preference. Here the authors assume that the GRS is a proxy for identifying unique populations.

- Environmental Representativeness Score (ERS): This score indicates the adequacy of environmental coverage of genebank accessions. The ERS is obtained by comparing the existing germplasm accessions with the complete environmental range of the ENM. The ENM is used as a proxy for abiotic adaptations, assuming that a different environment has an associated adapted population responding to its particular environmental conditions.

These three metrics (SRS, GRS, ERS) are combined in an equally-weighted average to produce a "Final Priority Score" with a scale between 0 to 10, where 0 is the minimum score indicating that the taxon is currently unrepresented in terms of number of samples, geographic extent and environmental coverage, while 10 is the maximum score indicating adequate representativeness of samples available on genebanks. A "Final Priority Category" is produced based on the quantitative "Final Priority Score", taxa with scores between 0 and 3 or no genebanks samples are categorized as 'High Priority Species" (HPS), the "Medium priority species" category is assigned for values between 3.01 and 5, the "Low priority species" category is given when the numeric score is between 5.01 and 7.5, and the "No further urgent conservation required" (NFCR) category is assigned when the final priority score is between 7.51 and 10 (Ramírez-Villegas *et al.*, 2010). This methodology was firstly applied for the CWR of beans (*Phaseolus* spp.), and the results obtained were evaluated and contrasted by an expert authority on the genus *Phaseolus*, obtaining a high correlation on the numeric priority scores, and detecting an overall underestimation of priorities for conservation of the Gap Analysis when compared to the expert's opinion.

Conservation assessments can be complemented with geographic analysis thanks to the growing availability of data containing explicit geographic references, as shown in Ramírez-Villegas *et al.* (2010). These geographic references include: countries administrative boundaries, climate layers, soils maps, digital elevation models, georeferenced species occurrence records. Likewise, the reduction of costs of hardware for processing and storing information, and the development of specialized software for processing and analyzing this type of information facilitates the application of conservation studies with a strong geographic component.

The importance of combining geographic analysis with conservation assessments relies in the possibility of identifying priority areas for conservation, but also helps planning work in the field by identifying the areas where target species are likely to be found. For the case of PGRFA, conservation assessments with a strong geographic component include: optimization of field collections by targeting specific areas (Jarvis *et al.*, 2005), more efficient germplasm collections using homogeneous ecogeographical units (Parra-Quijano *et al.*, 2011), and acquisition of materials with potential adaptations to climatic extremes (Afonin and Greene, 1999).

1.6 Aims of the study

This thesis aims to identify crop wild relative taxa that are priorities for immediate conservation action, and the regions where such conservation actions need to be directed. Species distribution models are used to determine the geographies where crop wild relatives are likely to be found. An *ex situ* gap analysis is then performed to understand how well conserved are crop wild relatives in genebanks, and thus determine collecting priorities. Projections of species distribution models into future climates are then used to assess the potential impacts of climate change on the distributions of crop wild relatives. A set of complementing metrics are calculated to prioritize groups of crop wild relatives aggregated per associated crop (i.e., crop gene pool). The following objectives were set to achieve this aim:

1. Create an electronic database containing georeferenced occurrence records of crop wild relative taxa. This database has a global coverage and contains information gathered from multiple data sources.

2. Model the potential distributions of crop wild relative taxa.

3. Assess the degree of completeness of crop wild relatives genebank collections by performing an *ex situ* gap analysis. Based on the results obtained, identify taxa and geographic regions that are urgent for collecting and subsequent *ex situ* conservation.

4. Evaluate the potential impacts of climate change on the environmental niches of crop wild relatives.

5. Estimate a set of complementing metrics that can be used as additional layers to further refine conservation priorities.

In order to achieve the objectives enumerated before, this thesis is structured as follows:

- Chapter 2 provides a description of the creation of the global occurrence database of crop wild relatives. This chapter reports the procedures applied to gather, organize, standardize and validate data. The database here presented contains data on reference records (e.g., botanical specimens, referenced sightings, unavailable germplasm accessions), and germplasm accessions. The occurrence data presented in this chapter are key inputs to produce the species distribution models that are used in Chapters 3, 4, 5, 6 and 7. Likewise, the information on germplasm accessions is used for the *ex situ* gap analyses performed in Chapters 3, 4 and 5.

- The potential distribution maps of crop wild relatives are produced by using the occurrence records stored in the global database presented in Chapter 2, and environmental layers as inputs of MaxEnt, a modelling algorithm. These potential distribution maps are in turn used in the *ex situ* gap analyses, and also used to produce richness maps indicating areas where distinct crop wild relative taxa may overlap.

- *Ex situ* gap analyses are initially performed with two individual case studies, presented in Chapter 3 and Chapter 4. The *ex situ* gap analysis consists of evaluating the representativeness of crop wild relatives in genebanks in terms of the gross number of accessions conserved in genebanks (compared to that existing as reference sightings), the proportion of the geographic extent (i.e., species potential distribution) represented as accessions in genebanks, and the proportion of distinct ecological units represented in genebanks. Each of these chapters explore in detail the how well represented are the crop wild relatives of two distinct associated crops: potato and eggplant, and conservation considerations particular to each case. Chapter 3 explores the *ex situ* conservation needs of the wild relatives of potato, a group of taxa with a long history of use in plant breeding. This chapter concludes with the identification of three levels of priority for further conservation, and observations on the urgency of conservation action because of the rapid land use changes experienced in the habitats of potato wild relatives. Chapter 4 is devoted to analyze the *ex situ* needs of conservation of the wild relatives of eggplant. This chapter includes a definition of the gene pool concept for eggplant built upon information on crop wild relatives and crop crossability and recent taxonomic revisions. This chapter reports an extensive analysis of different conservation aspects of the crop wild relatives of eggplant, including a prioritization of taxa and areas for *ex situ* and *in situ* conservation.

- An *ex situ* gap analysis performed in 81 different crop gene pools is presented in Chapter 5. This chapter summarizes the analyses of potato (Chapter 3), earlier versions of the eggplant analysis (Chapter 4) and 79 additional crop gene pools. Chapter 5 analysis to what extent the potential diversity of 1,076 crop wild relatives (associated to 81 crop gene pools) is available through genebanks, which taxa are considered as high priorities for urgent collecting and *ex situ* conservation because of their low extent of representativeness, and the regions considered as collecting hotsposts, where multiple high priority taxa for collecting are likely to occur. In addition, this chapter presents a set of global priorities for crop wild relatives.

- Chapters 6 and 7 provide further elements that can help refining the conservation priorities presented in 5. Chapter 6 discusses the potential effects of climate change on the distributions of crop wild relatives. This chapter identifies the crop gene pools and individual taxa that are most susceptible to climate change, the susceptible taxa that are also poorly represented in genebanks (as per the results presented in Chapter 5), and the crop gene pools and taxa that may not be largely affected under a future scenario of climate change.

- Chapter 7 focuses on producing a set of complementary metrics that can be used as additional layers of analysis to further refine conservation priorities (Chapter 5). This chapter uses publicly available statistics of global agricultural production, global contributions of macronutrients for diets, and the likelihood of using crop wild relatives in plant breeding, to produce three complementary metrics for refining conservation priorities. These metrics are used to estimate the value of crop gene pools, and in consequence the value of crop wild relatives. The chapter concludes with the identification of crops and regions that are of highest priority given the value of the crop (and therefore crop wild relatives) and the urgency for collecting (Chapter 5).

The work presented in this thesis is expected to provide an overview of the needs of *ex situ* conservation of the crop wild relatives of multiple crops. It is also expected that the methodologies, data and approaches used in this thesis can be used to analyze the conservation needs of individual crop wild relative taxa, other crop gene pools not included in this study, and different spatial scales (e.g., global, regional, national and/or subnational).

CHAPTER 2

A GLOBAL OCCURRENCE DATASET FOR CROP WILD RELATIVES

The work presented in this chapter is being submitted to Scientific Data.

Nora P. Castañeda-Álvarez, Colin K. Khoury, Chrystian C. Sosa, Steven Sotelo, Ingrid Vanegas, Shirley S. Calderón, Holly Vincent, Ruth H. Harker, Vivian Bernau, Ruth J. Eastwood, Julian Ramirez-Villegas, Hannes Dempewolf, Luigi Guarino, Jane Toll, Andy Jarvis, Jonas Müller, the Global Consortium of Crop Wild Relative Occurrence Data Providers* and Nigel Maxted.

N.P.C-A., C.K.K., H.D., R.J.E., L.G., R.H.H., A.J., N.M., J.M., J.R-V., H.V. and J.T. conceived the study, searched for and coordinated the collection of occurrence data. N.P.C-A., C.K.K., R.H.H. and H.V. collected occurrence data. N.P.C-A., C.K.K., J.R-V., S.H.S., H.V., S.K., R.J.E, and N.M. designed the database. S.H.S. developed and documented the standalone tool in java. I.V., S.C. and C.S.S. digitized data from herbaria and gray literature. C.S.S., C.K.K. and V.B. standardized taxonomic names. H.A. analyzed the survey responses from experts. N.M. and the Global Consortium of Crop Wild Relative Occurrence Data Providers contributed occurrence data. N.P.C-A led the writing with major inputs from C.K.K. and N.M., and further inputs from all other authors.

* For the complete list of authors please refer to the Supplementary Table 2.1

2.1 Abstract

The wild relatives of cultivated plants are important sources of traits for agriculture thanks to their rich genetic diversity. For being used, crop wild relatives (CWR) taxa and populations must be conserved and available for plant breeders via germplasm banks, however many CWR taxa are currently underrepresented in these facilities. Furthermore, complementary *in situ* conservation efforts currently cover few CWR taxa and populations. Improving the representativeness of CWR taxa in germplasm banks, and defining new sites for *in situ* conservation can be achieved by prioritizing CWR taxa and geographical sites. In this context, databases storing information on the taxa identity, and detailed descriptions of the localities where the taxa have been recorded are critical to understand the distributions of species, detect hotspots of taxa richness, and assess threats that may affect the habitats of CWR. To support these needs, an open-access global database was created by assembling occurrence records obtained from digital databases, literature, researchers' archives and herbaria. The database provides information on the geographic distributions of CWR. All records were subjected to standardization, and validation to secure the quality of the information stored in the database. This dataset provides a valuable source for researchers in the fields of conservation, population ecology, plant diversity, evolution, domestication, and rehabilitation of ecosystems.

2.2 Background and summary

Agriculture is challenged to continue producing food for a growing population under a changing climate (Porter *et al.*, 2014). Furthermore, natural resources required to produce more food are projected to become scarcer. In the same way the negative impacts of agriculture to the environment need to be reduced (Rosegrant and Cai, 2001; Van Vuuren *et al.*, 2010; Johnson *et al.*, 2014). Sustainable intensification of food production proposes various mechanisms to respond to these challenges, including the development of more productive and better adapted varieties through plant breeding (Foley *et al.*, 2011). Plant genetic resources are used as the main sources of genetic diversity and agronomic traits for developing new crop varieties. Crop wild relatives, a class of plant genetic resources, play a crucial role in plant breeding, particularly for pest and disease resistance, tolerance to abiotic stresses, quality and nutritional traits (Abberton *et al.*, 2015). Their use is increasing due to the advances in breeding technologies and lack of desired traits within the crop itself, but is constrained by the limited availability of these wild genetic resources in genebanks and general lack of information on species identities and distributions (Tanksley and McCouch, 1997; FAO, 2010; Khoury *et al.*, 2015a; Castañeda-Álvarez *et al.*, 2016). More than 70% of the wild relatives associated to crops of global importance are currently either unrepresented or underrepresented in genebanks (Castañeda-Álvarez *et al.*, 2016). Moreover, crop wild relatives are threatened in the wild by habitat destruction, invasive species, and climate change (Jarvis *et al.*, 2008; Kell *et al.*, 2012; Ureta *et al.*, 2012; Brummitt *et al.*, 2015b).

Efforts to enhance the conservation, availability and use of crop wild relatives are taking place by collecting, conserving and preparing the crop wild relatives of more than 20 crops important for global food security, like cereals, roots and tubers (Dempewolf *et al.*, 2013). Complementary initiatives are still needed to conserve crop wild relatives in their habitats (*in situ* conservation), and to further increase the representativeness in genebanks of the wild relatives of vegetables and fruits. In terms of facilitating access to crop wild relative plant material (and plant genetic resources in general), genebanks are setting-up digital catalogues where seeds can be requested, and the associated passport data of each germplasm accession are being made publicly available.

Genebank databases storing the passport data of germplasm accessions, and herbaria databases holding the voucher information of botanical specimens are the main sources of occurrence data of crop wild relative taxa. Both types of database contain information on the identity of each sample, descriptions of the locality where the sample was taken (more recent collected samples contain geographic coordinates), the collector(s), sample identifiers, common names, reported local uses, descriptions of the habitat where the sample was acquired, and institutions holding duplicates of the material. As the objectives of collecting germplasm differ from those guiding a botanical expedition, different types of information are found in germplasm databases when compared to herbaria databases. For instance, germplasm databases contain information on the biological status of the sample (i.e., whether it corresponds to wild, weedy, traditional cultivar, landrace, breeding or research material, advanced or improved cultivar, or GMO), and the type of germplasm storage (e.g., seed collection, field collection, *in vitro* collection, cryopreserved collection or DNA collection). While herbaria databases provide information on who did the taxonomic identification of the specimen, when was it identified, taxonomic determinations available on the specimen, the type status of the specimen (e.g., holotype, lectotype, isotype, paratype, neotype), and whether the plant had flowers, fruits or seeds at the moment of being collected.

Currently, many genebanks and herbaria have embarked on the digitization and mobilization of information related to accessions and botanical specimens held in their collections. However, large collections remain in non-digitized formats, or have applied different standards that prevent the integration with larger existing databases. In the case of germplasm accessions, there are about 450 institutes holding germplasm collections that have made their collections digitally available, and thus visible to the global community, however there are about 1,400 that still need to prepare their collections to be digitally available. In terms of botanical specimens, herbaria are digitizing their collections under the umbrella of various initiatives (e.g., Global Plants Initiative), and the mobilization of digital records is increasingly growing thanks to initiatives like the Global Biodiversity Information Facility, which offers access to 172 million Plant records (September 2015). Previous compilations of crop wild relatives' occurrence records are common within the frame of individual projects, however to the date, none of them have provided web-based and open-access to the complete set of occurrence records utilized for their analyses and studies.

Here we present a comprehensive description of the process followed to develop and validate the global occurrence dataset of crop wild relatives. This is the first database devoted to the compilation of information on the taxa names and localities of crop wild relatives in the world. The database has been populated with information derived from digital portals, herbaria, researchers' archives and literature. The database stores information on species and where they have been found in the world (at national, and sub-national level). The database is free of access here: `http://www.cwrdiversity.org/checklist/cwr-occurrences.php`. Both the structure and technical validation are described to allow updates of the existing database and compatibility with subsequent initiatives..

2.3 Methods

Close collaborations were established with researchers and institutions, who have accumulated a wealth of occurrence records, to gather the data making up the global occurrence dataset for crop wild relatives (Supplementary Table 2.2). After the data collection phase was completed, all these data needed to be standardized and curated in order to develop a single unified database containing all the relevant information associated with crop wild relatives records. The following section describes in detail the process that was used to gather and prepare the occurrence records that conform the global database. A summarized scheme of these processes is shown in Figure 2.1.

2.3.1 Data collection

Here we provide additional detail on the methodology of data collection used in Castañeda-Álvarez *et al.* (2015, 2016); Khoury *et al.* (2015a,b); Kantar *et al.* (2015), which utilize subsets of the dataset presented here.

An initial list of 193 genera belonging to 44 different botanical families and related to more than 130 crops of global importance was prepared to guide the collection of occurrence data. Commonly, a single genus is associated to a cultivated plant species, but for some specific cases (e.g., maize and wheat), more than one genus is associated to the crop, thus these additional genera were also included in the initial list (Table 2.1). The objective of gathering information for all members of a genus of crops is to collect information of crop wild relatives of present

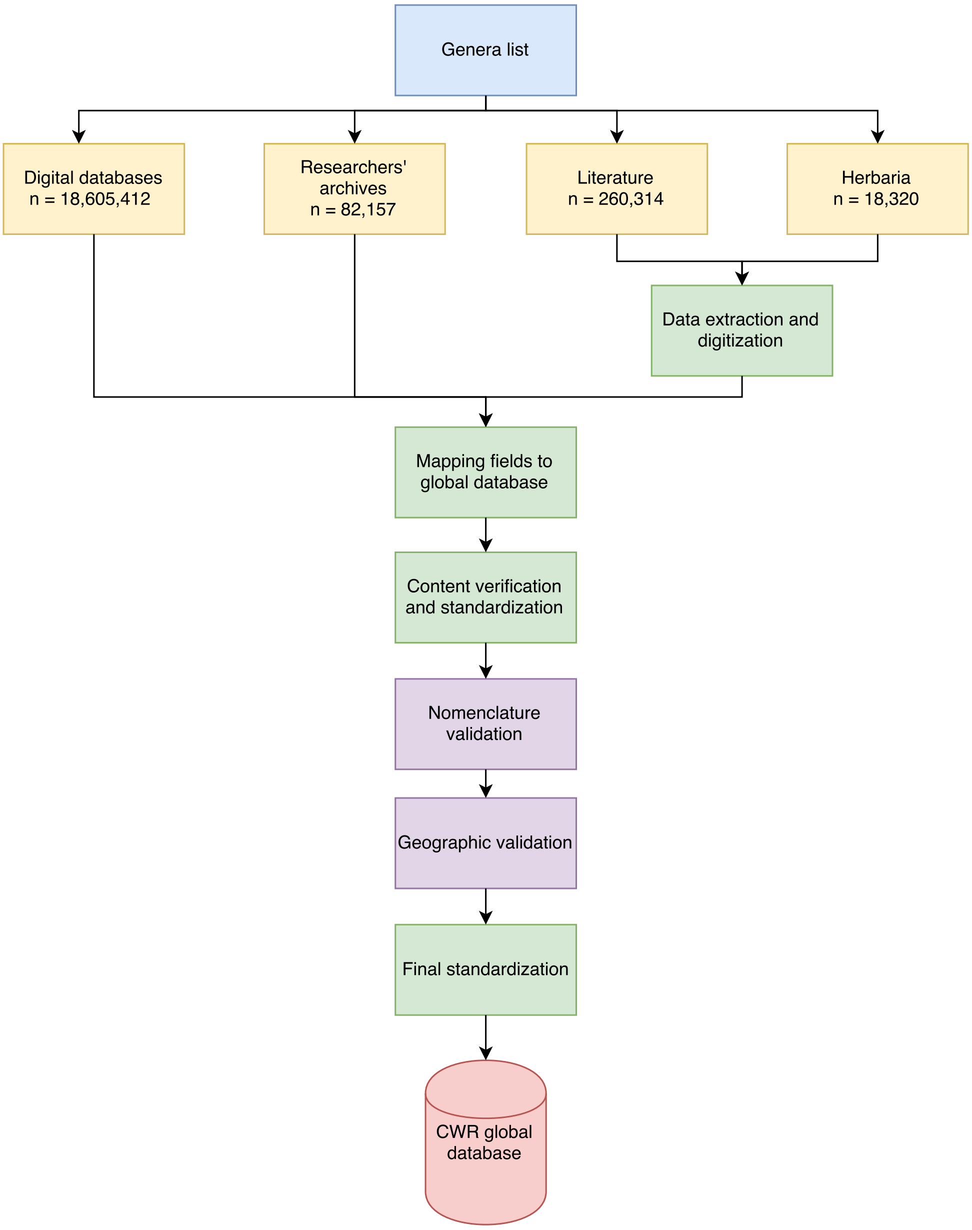


Figure 2.1: Scheme of the process of collecting, preparing and validation crop wild relatives' occurrence data. The blue box corresponds to the list that was used as reference for gathering occurrence data; yellow boxes correspond to the major groups of data sources; green boxes represent processes; purple boxes represent validation steps; and the red cylinder represents the final database.

and future use for plant breeding. Occurrence data were obtained from 420 data sources. We categorized these data sources in four major groups: digital databases, researchers, literature, and herbaria.

Table 2.1: List of genera and associated crop name used for guiding the collection of occurrence data

Family	**Genus (Crop name)**	**Family**	**Genus (Crop name)**	**Family**	**Genus (Crop name)**	**Family**	**Genus (Crop name)**
Sapindaceae	*Acer* (Sugar maple)	Rubiaceae	*Coffea* (Coffee)	Schisandraceae	*Illicium* (Badian)	Poaceae	*Psathyrostachys* (Wheat)
Poaceae	*Aegilemma* (Wheat)	Brassicaceae	*Coincya* (Cabbage, horseradish, rapeseed, turnip)	Poaceae	*Imperata* (Sugarcane)	Poaceae	*Pseudoroegneria* (Wheat)
Poaceae	*Aegilonearum* (Wheat)	Araceae	*Colocasia* (Yautia)	Convolvulaceae	*Ipomoea* (Sweetpotato)	Myrtaceae	*Psidium* (Guava)
Poaceae	*Aegilopodes* (Wheat)	Rosaceae	*Comarum* (Strawberry)	Brassicaceae	*Isatis* (Cabbage, horseradish, rapeseed, turnip)	Rosaceae	*Pyrus* (Pear)
Poaceae	*Aegilops* (Wheat)	Poaceae	*Comopyrum* (Wheat)	Convolvulaceae	*Jacquemontia* (Sweetpotato)	Brassicaceae	*Raphanobrassica* (Cabbage, horseradish, rapeseed, turnip)
Poaceae	*Agropyron* (Wheat)	Convolvulaceae	*Convolvulus* (Sweetpotato)	Juglandaceae	*Juglans* (Walnut)	Brassicaceae	*Raphanus* (Cabbage, horseradish, rapeseed, turnip, radish)
Amaryllidaceae	*Allium* (Garlic, onion and leek)	Betulaceae	*Corylus* (Hazelnut)	Poaceae	*Kiharapyrum* (Wheat)	Poaceae	*Rhaphis* (Radish)
Poaceae	*Alopecurus* (Sugarcane)	Brassicaceae	*Crambe* (Cabbage, horseradish, rapeseed, turnip)	Fabaceae	*Lablab* (Lablab)	Fabaceae	*Rhynchosia* (Pigeonpea)
Poaceae	*Amblyopyrum* (Wheat)	Poaceae	*Critesion* (Barley)	Asteraceae	*Lactuca* (Lettuce)	Grossulariaceae	*Ribes* (Currant and gooseberry)
Vitaceae	*Ampelocissus* (Grape)	Poaceae	*Crithodium* (Wheat)	Fabaceae	*Lathyrus* (Grasspea)	Brassicaceae	*Rorippa* (Cabbage, horseradish, rapeseed, turnip)
Rosaceae	*Amygdalus* (Plums, apricot)	Cucurbitaceae	*Cucumis* (Melon)	Fabaceae	*Lens* (Lentil)	Rosaceae	*Rubus* (Raspberry and blackberry)
Bromeliaceae	*Ananas* (Pineapple)	Cucurbitaceae	*Cucurbita* (Pumpkin and squash)	Fabaceae	*Lentilla* (Lentil)	Asteraceae	*Rudbeckia* (Rudbeckia)
Poaceae	*Andropogon* (Sorghum)	Rosaceae	*Cydonia* (Quince)	Brassicaceae	*Lepidium* (Cress)	Poaceae	*Saccharum* (Sugarcane)
Fabaceae	*Arachis* (Peanut)	Poaceae	*Cylindropyrum* (Wheat)	Poaceae	*Leymus* (Wheat)	Poaceae	*Sarga* (Sorghum)
Rosaceae	*Armeniaca* (Apricot)	Asteraceae	*Cynara* (Artichoke)	Fabaceae	*Lupinus* (Lupinus)	Poaceae	*Sclerostachya* (Sugarcane)
Brassicaceae	*Armoracia* (Cabbage, horseradish, rapeseed, turnip)	Poaceae	*Cynosurus* (Finger millet)	Solanaceae	*Lycopersicon* (Tomato)	Fabaceae	*Scytalis* (Cowpea)
Moraceae	*Artocarpus* (Breadfruit)	Fabaceae	*Cytisus* (Common broom, hairy-fruited broom, white spanish broom)	Rosaceae	*Malus* (Apple)	Poaceae	*Secale* (Rye)
Asparagaceae	*Asparagus* (Asparagus)	Apiaceae	*Daucus* (Carrot)	Anacardiaceae	*Mangifera* (Mango)	Pedaliaceae	*Sesamum* (Sesame)
Rutaceae	*Atalantia* (Orange)	Poaceae	*Digitaria* (Fonio)	Euphorbiaceae	*Manihot* (Cassava)	Poaceae	*Setaria* (Foxtail millet)
Fabaceae	*Atylosia* (Pigeonpea)	Dioscoreaceae	*Dioscorea* (Yam)	Fabaceae	*Medica* (Alfalfa)	Brassicaceae	*Sinapis* (Mustard seed)
Poaceae	*Avena* (Oat)	Ebenaceae	*Diospyros* (Persimmon)	Fabaceae	*Medicago* (Alfalfa)	Poaceae	*Sitopsis* (Wheat)
Fabaceae	*Azukia* (Adzuki bean)	Brassicaceae	*Diplotaxis* (Cabbage, horseradish, rapeseed, turnip)	Poaceae	*Milium* (American milletgrass)	Solanaceae	*Solanum* (Tomato, potato and eggplant)
Brassicaceae	*Barbarea* (Water cress)	Fabaceae	*Dolichos* (Lablab)	Poaceae	*Miscanthus* (Sugarcane)	Poaceae	*Sorghum* (Sorghum)
Lauraceae	*Beilschmiedia* (Avocado)	Poaceae	*Echinochloa* (Barnyard millet, japanese millet)	Brassicaceae	*Moricandia* (Cabbage, horseradish, rapeseed, turnip)	Amaranthaceae	*Spinacia* (Spinach)
Lecythidaceae	*Bertholletia* (Brazil nut tree)	Arecaceae	*Elaeis* (Oil palm)	Asteraceae	*Mulgedium* (Lettuce)	Myrtaceae	*Syzygium* (Clove)
Amaranthaceae	*Beta* (Sugar beet, beet and chard)	Zingiberaceae	*Elettaria* (Cardamom)	Musaceae	*Musa* (Bananas and plantains)	Malvaceae	*Theobroma* (Cacao)
Brassicaceae	*Brassica* (Cabbage, horseradish, rapeseed, turnip)	Poaceae	*Eleusine* (Finger millet)	Solanaceae	*Nicotiana* (Tobacco)	Malvaceae	*Thespesia* (Cottonseed)
Fabaceae	*Cajanus* (Pigeonpea)	Poaceae	*Elymus* (Wheat)	Poaceae	*Ochthochloa* (Finger millet)	Poaceae	*Thinopyrum* (Wheat)
Theaceae	*Camellia* (Tea)	Rosaceae	*Emplectocladus* (Ethiopian cabbage, cabbage, black mustard, mustard, rape)	Oleaceae	*Olea* (Olive)	Apiaceae	*Tornabenea* (Carrot)
Fabaceae	*Cantharospermum* (Pigeonpea)	Musaceae	*Ensete* (Banana)	Fabaceae	*Ononis* (Restharrow)	Brassicaceae	*Trachystoma* (Mustard seed)
Brassicaceae	*Capsella* (Cabbage, horseradish, rapeseed, turnip)	Brassicaceae	*Eruca* (Cabbage, horseradish, rapeseed, turnip)	Poaceae	*Orrhopygium* (Wheat)	Fabaceae	*Trigonella* (Fenugreek seed)
Solanaceae	*Capsicum* (Chillies)	Brassicaceae	*Erucastrum* (Cabbage, horseradish, rapeseed, turnip)	Brassicaceae	*Orychophragmus* (Cabbage, horseradish, rapeseed, turnip)	Poaceae	*Tripsacum* (Maize)
Caricaceae	*Carica* (Papaya)	Fabaceae	*Ervum* (Lentil)	Poaceae	*Oryza* (Rice)	Poaceae	*Triticosecale* (Triticale)
Asteraceae	*Carthamus* (Safflower)	Fabaceae	*Faba* (Faba bean)	Poaceae	*Padia* (Rice)	Poaceae	*Triticum* (Wheat)

Table 2.1 – continued from previous page

Family	Genus (Crop name)	Family	Genus (Crop name)	Family	Genus (Crop name)	Family	Genus (Crop name)
Juglandaceae	*Carya* (Pecan)	Moraceae	*Ficus* (Fig)	Rosaceae	*Padus* (Cherry)	Ericaceae	*Vaccinium* (Blueberry and cranberry)
Fagaceae	*Castanea* (Chestnut)	Apiaceae	*Foeniculum* (Fennel)	Poaceae	*Panicum* (Little millet, proso millet)	Orchidaceae	*Vanilla* (Vanilla)
Rosaceae	*Cerasus* (Stone fruits)	Rutaceae	*Fortunella* (Orange, lemon and grapefruit)	Amaranthaceae	*Patellifolia* (Sugar beet, beet and chard)	Caricaceae	*Vasconcellea* (Papaya)
Poaceae	*Chaetochloa* (Foxtail millet)	Rosaceae	*Fragaria* (Strawberry)	Poaceae	*Patropyrum* (Wheat)	Fabaceae	*Vavilovia* (Pea)
Poaceae	*Chamaeraphis* (Foxtail millet)	Poaceae	*Gastropyrum* (Wheat)	Poaceae	*Pennisetum* (Pearl millet)	Euphorbiaceae	*Vernicia* (Tung nut)
Arecaceae	*Chamaerops* (Date palm)	Poaceae	*Gigachilon* (Wheat)	Lauraceae	*Persea* (Avocado)	Fabaceae	*Vicia* (Faba bean and vetch)
Poaceae	*Chennapyrum* (Wheat)	Fabaceae	*Glycine* (Soybean)	Fabaceae	*Phaseolus* (Bean)	Fabaceae	*Vigna* (Cowpea)
Amaranthaceae	*Chenopodium* (Quinoa)	Malvaceae	*Gossypium* (Cottonseed)	Arecaceae	*Phoenix* (Date palm)	Sapotaceae	*Vitellaria* (Karite nut, shea nut)
Fabaceae	*Cicer* (Chickpea)	Grossulariaceae	*Grossularia* (Gooseberry, currant)	Myrtaceae	*Pimenta* (Pimento)	Vitaceae	*Vitis* (Grape)
Lauraceae	*Cinnamomum* (Cinnamon)	Asteraceae	*Harpalium* (Sunflower)	Piperaceae	*Piper* (Pepper)	Fabaceae	*Voandzeia* (Mustard seed)
Vitaceae	*Cissus* (Grape)	Asteraceae	*Helianthus* (Sunflower)	Anacardiaceae	*Pistacia* (Pistachio)	Araceae	*Xanthosoma* (Yams)
Cucurbitaceae	*Citrullus* (Watermelon)	Brassicaceae	*Hirschfeldia* (Cabbage, horseradish, rapeseed, turnip)	Fabaceae	*Pisum* (Pea)	Poaceae	*Zea* (Maize)

Digital databases

Online and open-access digital databases are platforms that facilitate access and discovery of large quantities of occurrence records associated to plant species, including crop wild relatives. In total, 3,296,401 records associated to germplasm accessions, and 14,679,011 reference records (e.g., specimens collected in botanical expeditions, non-viable germplasm seed accessions, and reported field observations) were acquired from web-based databases. The Global Biodiversity Information Facility (GBIF; `www.gbif.org`), a platform providing extensive access to global biodiversity related data, was identified as the largest source of occurrence records for the genera prioritized (more than 15 million records). Semi-automatized protocols were used to obtain batches of approximately 92,000 records from GBIF by using the R package rgbif (Chamberlain *et al.*, 2015).

At the time of the data gathering phase (year 2011), the CGIAR's System-wide Information Network for Genetic Resources (SINGER), the European Plant Genetic Resources Catalogue (EURISCO), and the United States Department of Agriculture's Genetic Resources Information Network (GRIN) were the largest existing platforms to facilitate access to genetic resources and associated passport data information. Nowadays, these sources are also provided through a single-point access portal: Genesys, a global portal to information on plant genetic resources (`www.genesys-pgr.org`). We contacted the database managers of each of these platforms for procuring the complete sets of data mobilized by their portals. Additional digital databases retrieved are listed in Supplementary Table 2.2.

Researchers

Researchers in the fields of botany, plant genetic resources and biodiversity conservation, and plant breeding, were approached as many of them have compiled through their years of practice large datasets containing detailed and accurate occurrence data in terms of nomenclature identity and geographic locality of the site where the sample was obtained or recorded. This information is rarely fully available to the public, occasionally it is available as subsets in peer-reviewed literature, or in few cases it has not been digitized. Twenty-four datasets from experts at 17 institutions were obtained containing a total of 82,157 records. Supplementary Table 2.2 contains the names and affiliation of all collaborators.

Literature

Peer-reviewed and gray literature (e.g., project reports and graduate theses) were also a relevant source of occurrence data. Finished projects and various universities with CWR research groups were approached to obtain access to the occurrence records that have been used to estimate the potential distribution of crop wild relatives within the geographic scope of their work, to plan conservation actions, and to assess the likeness of crop wild relatives to cross with their associated crops. In total, 260,314 records were retrieved from finalized projects, literature, and academic dissertations.

Herbaria

A query to GBIF on the occurrence data available for flowering plants (Magnoliophyta) in October 15th, 2010 enabled us to identify countries and regions presenting an overall low density of georeferenced occurrence records (< 999 georeferenced records/per degree cell). Based on this assessment, it was possible to identify the countries with less information mobilized at that time through GBIF (Table 2.2). Using this list of countries, a query to the Index Herbariorum (Thiers, 2010) identified the herbaria holding large collections of botanical specimens collected in the countries with comparatively less information available through GBIF. A protocol prepared to capture data from herbaria allowed to collect 55,977 photographs from 22 herbaria. The botanical specimens selected for being photographed must have had geographical information in their voucher (e.g., geographical coordinates or detailed locality descriptions that would enable subsequent georeferencing), recent taxonomical determinations, and presence of reproductive structures. Two photographs were taken from each botanical specimen, a first photograph was taken to the overall specimen, accompanied with a paper note indicating whether the specimen had flowers or fruits, and a second picture was taken of the label data. All photographs were taken with cameras with a minimum resolution of five megapixels, as this eased the reading and subsequent digitization of the text in the specimen. All photographs were organized in folders according to the herbarium where they were obtained, and sub-folders according to the family and genus that the specimen belong to. These photographs are available upon request. A total of 18,320 records were obtained from visits to herbaria.

Table 2.2: Countries with low density of Magnoliophyta records mobilized through GBIF.org in October 2010

Country	**Continent**	**Herbarium (potentially holding information for each country)**
Brazil	America	COI, LISC
Chile	America	MA
Uruguay	America	MA
Argentina	America	MA
Cuba	America	MA
Mozambique	Africa	COI, LISC
Angola	Africa	COI, LISC
Zambia	Africa	K, E, BM, MANCH
Congo (DRC)	Africa	BR, BRLU
Congo	Africa	P
Central African Republic	Africa	P
Chad	Africa	P
Niger	Africa	P
Mauritania	Africa	P
Estonia	Europe	LE, MW, LECB, WIR
Latvia	Europe	LE, MW, LECB, WIR
Lituania	Europe	LE, MW, LECB, WIR
Rusia	Europe	LE, MW, LECB, WIR
Byelarus	Europe	LE, MW, LECB, WIR
Ucrania	Europe	LE, MW, LECB, WIR
Syria	Asia	P
Iraq	Asia	K, E, BM, MANCH
Iran	Asia	K, E, BM, MANCH, LE, MW, LECB, WIR
Georgia	Europe	LE, MW, LECB, WIR
Kazakhstan	Asia	LE, MW, LECB, WIR
Turkmenistan	Asia	LE, MW, LECB, WIR
Uzbekistan	Asia	LE, MW, LECB, WIR
Tajikistan	Asia	LE, MW, LECB, WIR
Kyrgyzstan	Asia	LE, MW, LECB, WIR
China	Asia	PE
Pakistan	Asia	K, E, BM, MANCH
India	Asia	K, E, BM, MANCH
Bangladesh	Asia	K, E, BM, MANCH
Myanmar (Burma)	Asia	K, E, BM, MANCH
Indonesia	Asia - Oceania	L, BO
Laos	Asia	P
Vietnam	Asia	P
Cambodia	Asia	P
North Korea	Asia	TI, LE, MW, LECB, WIR
Mongolia	Asia	LE, MW, LECB, WIR
Philippines	Asia	MA, NY, MO, HUH, US, A, F
Vanuatu	Oceania	P, K, E, BM, MANCH
Solomon Islands	Oceania	K, E, BM, MANCH

2.3.2 Data preparation

Data extraction and digitization

Data in non-digital and/or manipulable formats (e.g., published papers, fieldwork notes, and photographs of botanical specimens) were manually digitized and stored in a temporary digital database in a spreadsheet application. A complete list of the fields included in the temporary digital database is presented in Supplementary Table 2.3.

During the digitization process particular attention was given to information related to the identity of the specimen (e.g., family, genus, and scientific name) of the specimen, at least the three most recent taxonomical determinations, its dates and corresponding authors, the geographical description where the sample was taken (e.g., country of collection, coordinates) and date in which the specimen was collected, the collector name, whether the specimen had flowers and/or fruits at the moment of collection, any codes that may help to identify the specimen in each herbarium, and whether copies of the specimen were sent to other herbaria. Having these information available in a digital format is valuable for various analysis like producing potential distribution models, and understanding the patterns of flowering and seed production of a specific plant species.

Fields mapping

Despite the existence of protocols designed to standardize taxon occurrence databases and to facilitate database intercommunication (e.g., the Multi-Crop Passport Descriptors for germplasm accessions (Alercia *et al.*, 2012), the Darwin Core standard for biological diversity, and the Access to Biological Collections Data standard), all datasets obtained in the data collection phase used different standards, and thus a thorough analysis to understand the content of the fields in the original datasets was performed. This enabled us to map the fields of the individual datasets to the fields in the global database.

Content verification and standardization

The global occurrence database on crop wild relatives was created with the purpose of holding information for the production of analyses that help understanding the degree of *ex situ* and *in situ* conservation of crop wild relatives globally. Therefore, various procedures were applied to verify, complete and standardize the content of fields holding critical information for the production of these conservation analyses, specifically for fields containing information on plant genetic resources and crop wild relatives, fields with information that facilitate the identification of the original sources of the record, and fields holding details on the geographic locality where a sample was taken.

First, records were flagged according to the type of sample they represented (i.e., whether it is a germplasm accession or a reference record). Germplasm passport data give account of information indicating if the accession is actively conserved in germplasm banks, the availability of plant material to the public, botanic identity of the accession, whether the accession correspond to cultivated, wild material or an hybrid, and in some cases, information on the seed viability. For reference records (e.g., botanical specimens, reported field observations, and unavailable germplasm accessions) serve to report localities where a species has been recorded and may have not been collected for being conserved in a genebank. Data records obtained from genebank databases were marked as germplasm records, unless the accession is unavailable, and thus was reclassified into a reference record. For the GBIF mediated data, the information on data providers and dataset names associated to each record were used to detect the original source of data, and thus to infer whether the original corresponds to germplasm accessions or reference sightings. The World Information and Early Warning System of Plant Genetic Resources for Food and Agriculture (WIEWS; `http://www.fao.org/wiews-archive/wiews.jsp`) was used to verify if a data source recorded in GBIF corresponds to a germplasm bank and thus classify the records as possible germplasm accessions. Moreover, further interactions with researchers during the preparation of individual studies of the wild relatives of potato (Castañeda-Álvarez *et al.*, 2015), sweetpotato (Khoury *et al.*, 2015a), pigeonpea (Khoury *et al.*, 2015b), sunflower (Kantar *et al.*, 2015) and eggplant (Syfert *et al.*, 2016), enabled us to detect non-viable accessions which were subsequently classified as reference records.

Fields that help to trace back the data provider and original source were also standardized and made mandatory as the information contained in these fields facilitates giving the appropriate acknowledgement and thus citations in the case of using individual datasets. These information is also useful to detect duplicates between different sets of data. The identifiers of each individual dataset (e.g., file name, name of digitizer) were included in the major database as a mean to trace the processing and quality of data.

Final standardization

After validating the botanical identity and geographic locality, additional information on the distributional range of species was assigned to each record (e.g., the record is located within the native, non-native, introduced, or naturalized range of the species). Various concepts have been proposed to define the nativeness of a species, with some authors supporting that native species are those that have been long-established (i.e., before the Neolithic) in the territories where they are currently found (Webb, 1985), and others proposing a non-dichotomic approach and suggesting multiple shades of nativeness (Usher, 2000). Despite the lack of consensus in the definition of species' nativeness and with the purpose of providing a category with the crop wild relatives records in our database, we queried the "Harlan and de Wet Crop Wild Relative Inventory" (Vincent *et al.*, 2013) using the validated scientific name and the country of sampling registered in the database to define the nativeness category of our records. The "Harlan and de Wet Crop Wild Relative Inventory" largely relies in the distributional range information of GRIN Taxonomy (`http://www.ars-grin.gov/~sbmljw/cgi-bin/taxcrop.pl?language=en`; USDA ARS National Genetic Resources Program, n.d.) and this in turn relies in literature and direct consultations to experts. When a georeferenced record is mapped within the limits of its reported native range it is classified as native, if not, it is classified as non-native. Other categories were considered (e.g., naturalized, introduced) if available through the inventory, or if they were already defined in the record. The definition of the distributional range of a species and identifying the record as possible native or non-native is helpful to detect probable native wild populations (i.e., those within the boundaries of the native distribution), and weedy populations or records of cultivated species (i.e., those records classified as non-native or introduced).

Finally, each dataset was assembled into the major database after applying all the processes described here, including the validation steps. Our final database contains 18,966,203 records, from which 13,452,398 contain validated coordinates (Fig. 2.2). Records without coordinates are stored and available in the database as some of them can serve to provide information on existing germplasm accessions and recorded sightings, to provide information on the flowering patterns of the species, and some cases contain the description of the localities where the specimen/sample was collected/recorded and can be re-processed for calculating their geographical coordinates.

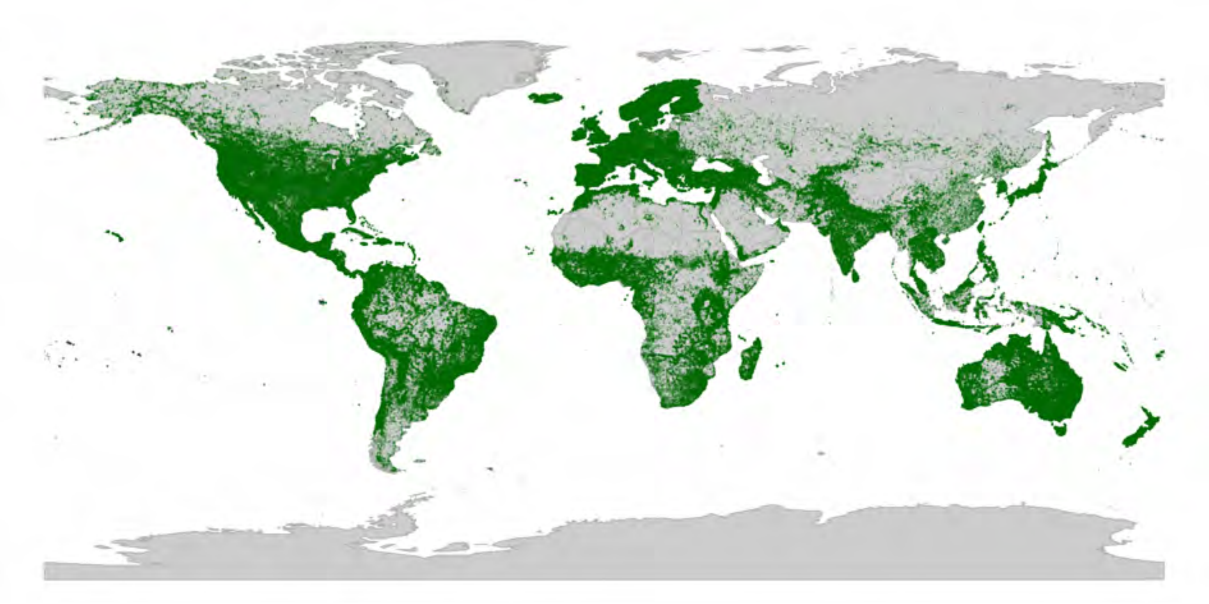


Figure 2.2: Distribution of occurrence records with geographic coordinates in the global database.

2.3.3 Code availability

The code prepared to obtain data from GBIF was prepared in R, and is available on GitHub `https://github.com/CIAT-DAPA/cwr_occurrencesvalidation/tree/master/RScripts`. The code designed to process and validate the occurrence records in terms of botanical identity and geo-spatial positioning consists of a stand-alone java application and is publicly available through GitHub (`https://github.com/CIAT-DAPA/cwr_occurrencesvali`

2.4 Data Records

For the complete list of data records see the Supplementary Table 2.4.

2.5 Technical Validation

We focus on validating the quality of the nomenclature identity and geospatial positioning of the occurrence records of crop wild relatives. This information is vital to circumscribe future analyses of the taxon(a) of interest for researchers, and enables the production of analyses with a component of spatial analyses. An automatized protocol was designed to validate and process each individual dataset, before being included in the global occurrence dataset of crop wild relatives. This protocol consists of two phases:

2.5.1 Nomenclature validation

In this step we verify that the scientific name (botanical identity) of each occurrence record does not contain misspellings and is standardized following accepted plant species checklists. For this, three tools providing plant species names are queried: GRIN (USDA ARS National Genetic Resources Program, n.d.), TNRS (Boyle *et al.*, 2013) and TaxonStand (Cayuela *et al.*, 2012). The names retrieved are then compared to the original scientific name of each record. Below the rules applied for defining the botanical identity of occurrence records:

- If the initial scientific name agreed with the name provided by GRIN, this name is assumed as the accepted name, as GRIN is frequently used as a standard for plant genetic resource nomenclature.

- If the initial scientific name is not found in GRIN, but coincides with the name provided by TNRS, the latter is used as the accepted name for the record.
- If the initial scientific name is not found in GRIN nor TNRS, but coincides with TaxonStand, then the name provided by TaxonStand is used.

When the scientific name does not coincide with none of the three sources of plant species names, a manual query is performed by checking JSTOR Global Plants (`https://plants.jstor.org/`), the International Plant Names Index (IPNI; *The International Plant Names Index*, 2012) and the Integrated Taxonomic Information System (ITIS; *Integrated Taxonomic Information System (ITIS)*, n.d.).

If the name is not found after the manual revision, the record is not included in the global database.

To summarize, 16,817,296 records are aligned to GRIN, 923,131 records are aligned to TNRS, 1,012,906 records are aligned to TaxonStand.

In terms of taxonomic resolution, all records contain information of their botanical family and genera, 18,215,161 have the specific epithet, and 1,301,651 records contain information at the subespecific level (subspecies and variety).

2.5.2 Geographic validation

Here we validate and standardize the content in the fields that contain information related to the geographic sites where the crop wild relatives' species have been recorded or sampled. Based on the original country of each record we standardize all the country names and corresponding ISO2 codes in the database. Likewise, the fields containing information of coordinates in the format DDMMSS (degree, minutes, seconds) and direction (North, South, West and East) are reviewed and standardized if required. Some records only had coordinates in the format DDMMSS, therefore it was necessary to calculate and populate the latitude and longitude fields with coordinates in decimal format.

Using the fields that describe the site of collection or recording (e.g., country, adm1, adm2, adm3, local_area and locality) an automatized query is initially performed in GEOLocate (`http://www.museum.tulane.edu/geolocate/`; Rios and Bart, 2010). Here some examples of the detailed locality descriptions that can be used for georeferencing:

- Burma, Kachin, Myitkyina, Chipwi, Sand bank of Nmai.
- Canada, Québec, Montréal, Parc Terra Cotta.
- Germany, Lower Saxony, Harburg, Salzhausen, Situated approximately 40 km southeast of Hamburg and 15 km west of Luneburg.

When no coordinates are obtained from this query, a second query is performed to the Google Maps Geocoding API. GEOLocate is preferred over the Google Maps Geocoding API to georeference species records as it uses natural history gazetteers for its processes, and its mean precision distance tends to be lower than the mean precision distance obtained with Google Maps Geocoding API (Fig. 2.3). Once the new geographic coordinates are estimated, only those coordinates with a precision distance below 5 km are considered for the subsequent validation steps, in agreement with the resolution used for the spatial modelling (∼5 x ∼5 sq km gridcell).

Subsequently, the coordinates originally provided with the species' records and the newly georeferenced coordinates with a precision distance below 5 km are compared with a raster (1 sq km resolution) to detect and remove coordinate points off land as in Warren *et al.* (2013). Afterwards, the remaining coordinates are checked whether they coincide with the geometric centroid of each country - a common error in batch-georeferencing - by comparing the values of the newly georeferenced coordinates against the coordinates of the geometric centroid. We flagged the records where the newly georeferenced coordinates coincide with the coordinates of the geometric centroid, and assessed in a record-by-record basis if it was a miscalculation of the georeferencing process (e.g., the country name was the only parameter used for georeferencing). All records with miscalculated centroid-coordinates were removed, as this may not represent the actual locality, habitat and environmental conditions where the taxon has been collected/reported, and thus there is a high risk of misguiding subsequent analyses.

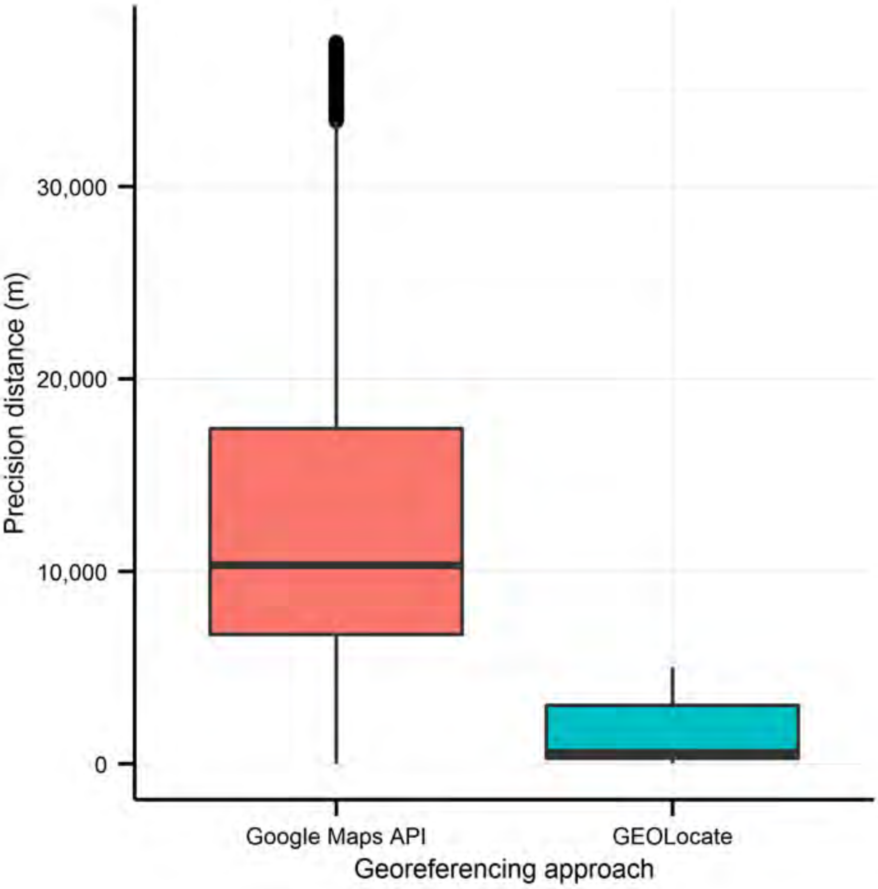


Figure 2.3: Boxplots of the precision distances of georeferenced coordinates using GEOLocate and Google Maps Geocoding API.

To detect conflicts between the country of collection and the coordinates describing the specific site of collection, we mapped the georeferenced occurrences on a global boundaries map and revised if the country name originally registered with the record coincides with the country where the coordinate is found, as described by Hijmans *et al.* (1999). The following rules were used:

- When the country name originally recorded with the record coincides with the country name where the occurrence record is mapped, the coordinate is accepted.
- If the country name originally recorded with the record does not coincide with the country name where the occurrence record is mapped, the coordinate is rejected, and the record is left without a geographic coordinate.

An additional validation was performed to a subset of data by presenting the occurrence records of the wild relatives of sixteen crops to 29 expert researchers. Researchers were asked to score the sets of occurrence records according to the completeness of geographic coverage and accuracy of the distributions of the CWR of the crops of their expertise. An overall agreement was found for most of the CWR, and less agreement in three crops (i.e., tomato, lettuce and oats) because the occurrence records presented did not covered the complete extension where the species have been collected and reported (Table 2.3).

Table 2.3: Experts' degree of agreement with the accuracy and completeness of the occurrence records of crop wild relatives.

Confidence in assessment of occurrence data	**Crop**
Strongly agree	Sorghum
Agree	Bananas and plantains, sugar beet, wheat, barley, rice, sweetpotato, pearl millet, spinach, pigeonpea
Neutral	Carrots, safflower, quinoa, oat
Disagree	Tomato, lettuce
Strongly disagree	

To summarize, 292,332 records initially plotted as occurring in the sea were corrected; 36,445 records were detected to coincide with country geometric centroids by performing a spatial analysis, and thus their coordinates were recalculated ; 1,848,020 records were re - georeferenced due to inconsistencies in their original geographic coordinates; 5,513,805 records do not contain coordinates; and 5,740,402 only have country information (no subnational or locality description that can enable georeferencing).

2.6 Usage Notes

The dataset presented here was initially developed for use in a global assessment of the extent of crop wild relatives diversity conserved in genebanks (Castañeda-Álvarez *et al.*, 2016). Individual crop studies have also benefited from this database: potato (Castañeda-Álvarez *et al.*, 2015), sweetpotato (Khoury *et al.*, 2015a), pigeonpea (Khoury *et al.*, 2015b), sunflower (Kantar *et al.*, 2015) and eggplant (Syfert *et al.*, 2016). Further regional analysis (e.g., European and the

Southern African Development Community), and a global *in situ* gap analysis are also being currently undertaken using the dataset.

The global occurrence dataset of crop wild relatives can be used to map the distributions of crop wild relatives, produce potential distribution maps based on the environmental drivers of each species, assess the likely impacts of climate change and other threats on the recorded populations of these species, detect populations within protected areas and other analysis related to the conservation of crop wild relatives. Likewise, a global occurrence dataset of crop wild relatives can provide inputs for baseline assessments of the progress of conservation actions, and can be a contribution towards the achievement of international targets on biodiversity knowledge (Aichi Target 19; Secretariat of the CBD, 2010).

In terms of species distribution modelling, this dataset can also be used as biased background points, which in turn helps to reduce the bias associated to the input points (Phillips *et al.*, 2009). This dataset can also be matched with information derived from genetic assessments to perform phytogeographic studies, analysis on the structure and ecology of crop wild relatives' populations, and evolutionary processes affecting crop wild relatives.

2.7 Discussion

Improving the availability of species occurrence data in digital and reusable formats is urgent, as these data are key for understanding the distributions of species, and subsequent analysis that can inform conservation efforts (Meyer *et al.*, 2015).

Efforts to digitize the information contained in botanical specimens (e.g., GBIF, and the Global Plants Initiative) and germplasm passport data (e.g., GRIN-Global) are expected to continue contributing to the growth of occurrence data of plants (including crop wild relatives) available in digital formats. These efforts might contribute to fill the gaps in information of some regions like the tropics (namely the Amazonian rainforest, Central Africa, the Malay archipelago) and Asia (Eastern, Northern and Central Asia specifically) (see Figure 2.2).

Plans to increase the visibility of the global occurrence dataset for crop wild relatives include its publication through GBIF as a separate collection of agrobiodiversity related data, and at the same time securing its availability in a stable data repository (e.g., The Dataverse Project).

Further efforts to enhance the quality of occurrence records are needed. This includes improving taxonomic coverage, record completeness, correction of geographic biases, and taxonomic identity of occurrence records (Tittensor *et al.*, 2014; Goodwin *et al.*, 2015). This in turn will help to improve the quality and accuracy of the studies derived from the use of this data. Our approach enabled us to identify and overcome quality issues related to geographical positioning and taxonomic identity. We were able to calculate geographic coordinates when sufficient locality descriptions where available, to re-calculate coordinates when geographic inconsistencies were found, to correct misspellings in the botanic identities of occurrence records and to align the botanic identities of occurrence records to accepted plant checklists. The continuous interaction and feedback from crop wild relatives' experts was key to improve the quality of the data for some crop wild relative genera.

CHAPTER 3

EX SITU CONSERVATION PRIORITIES FOR THE WILD RELATIVES OF POTATO (*SOLANUM* L. SECTION *PETOTA*)

Article published in PLoS ONE 2015. DOI: `10.1371/journal.pone.0122599`

Nora P. Castañeda-Álvarez, Stef de Haan, Henry Juárez, Colin K. Khoury, Harold A. Achicanoy, Chrystian C. Sosa, Vivian Bernau, Alberto Salas, Bettina Heider, Reinhard Simon, Nigel Maxted and David M. Spooner.

NPC-A, SdH, CKK, HJ, and NM conceived and designed the study. NPC-A, HJ, CKK, HAA, CCS, and VB acquired data and performed the analyses. NPC-A, SdH, HJ, CKK, HAA, CCS, VB, AS, BH, RS, NM and DMS analyzed the results. NPC-A and SdH wrote the manuscript. NPC-A, SdH, HJ, CKK, AS, BH, RS, NM and DMS edited the manuscript.

3.1 Abstract

Crop wild relatives have a long history of use in potato breeding, particularly for pest and disease resistance, and are expected to be increasingly used in the search for tolerance to biotic and abiotic stresses. Their current and future use in crop improvement is dependent upon their availability in *ex situ* germplasm collections. As these plants are impacted in the wild by habitat destruction and climate change, actions to ensure their conservation *ex situ* become ever more urgent.

We analyzed the state of *ex situ* conservation of 73 of the closest wild relatives of potato (*Solanum* section *Petota*), with the aim of establishing priorities for further collecting to fill important gaps in germplasm collections. A total of 32 species (43.8%), were assigned high priority for further collecting due to severe gaps in their *ex situ* collections. Such gaps are most pronounced in the geographic center of diversity of the wild relatives in Peru. A total of 20 and 18 species were assessed as medium and low priority for further collecting, respectively, with only three species determined to be sufficiently represented currently. Priorities for further collecting include: (i) species completely lacking representation in germplasm collections; (ii) other high priority taxa, with geographic emphasis on the center of species diversity; (iii) medium priority species. Such collecting efforts combined with further emphasis on improving *ex situ* conservation technologies and methods, performing genotypic and phenotypic characterization of wild relative diversity, monitoring wild populations *in situ*, and making conserved wild relatives and their associated data accessible to the global research community, represent key steps in ensuring the long-term availability of the wild genetic resources of this important crop.

3.2 Introduction

Potato (*Solanum tuberosum* L.) is the most important tuber crop worldwide, continuing to gain significance in temperate and tropical regions as a source of carbohydrates, vitamins, and minerals (Khoury *et al.*, 2014) as well as for industrial purposes (Kraak, 1992). The crop is susceptible to a wide range of biotic stresses, in particular fungal diseases and pests (Stevenson *et al.*, 2001; Wale *et al.*, 2008). A relatively low historical influx of variation has led to a genetic bottleneck within potato cultivars (Mendoza and Haynes, 1974; Hawkes, 1979; Wang *et al.*, 2011), thus the infusion of potato varieties with novel genetic diversity is expected to improve resistance to biotic and abiotic stresses (Jansky *et al.*, 2013).

As one source of such variation, potato breeding programs have looked to related wild species, particularly for resistance to biotic and abiotic stresses (Rudorf, 1958; Pavek and Corsini, 2001; Jansky *et al.*, 2013). Widely used and well documented sources of valuable traits such as frost and late blight (*Phytophthora infestans* (Mont.) de Bary) resistance include *S. acaule*, *S. bulbocastanum*, *S. chacoense*, *S. demissum* and *S. stoloniferum*. The search for late blight resistance has been a center point in the evaluation and use of wild relatives in potato breeding (Mattheij *et al.*, 1992; van der Vossen *et al.*, 2003; Liu and Halterman, 2009; Smyda *et al.*, 2013; Jones *et al.*, 2014). In addition, *S. commersonii* and *S. berthaultii* have been evaluated for bacterial wilt (*Ralstonia solanacearum* Smith) and verticillium wilt (*Verticillium* spp.) resistances, respectively (Laferriere *et al.*, 1999; Hijmans *et al.*, 2003; González *et al.*, 2013). Other species have been proposed as valuable sources of resistance, such as *S. acroglossum* for Colorado potato beetle (*Leptinotarsa decemlineata* Say), and *S. albicans* for cold sweetening (Pelletier, 2007; Luthra *et al.*, 2009) (Table 3.1).

Despite the extensive history of use of the wild relatives of potato in breeding, most species have not yet been evaluated for their potential for utilization. These include species from the eastern Andean slopes where resistance to late blight is particularly key for survival (e.g., *S. laxissimum* and *S. rhomboideilanceolatum*), as well as distant relatives that may display drought resistance due to their adaptation to dry habitats (e.g., *S. immite* and *S. mochiquense*). Enhanced understanding of species reproductive biologies, advances in pre-breeding technologies to bypass reproductive barriers, improvements in cisgenic techniques, and the evolution of new genotyping and phenotyping platforms are likely make the use of wild relatives more attractive and efficient (Watanabe *et al.*, 1995; Hanneman, 1999; Camadro *et al.*, 2004; Jansky *et al.*, 2006;

Table 3.1: Crop wild relatives that have been evaluated and/or used in potato breeding.

Gene pool	**Species**	**Resistance trait(s)**	**References**
Primary	*S. acaule*	Biotic: *Nacobbus aberrans*. Abiotic: frost	(Estrada, 1980; Ross, 1986; Suárez *et al.*, 2009; Watanabe *et al.*, 1994)
	S. berthaultii	Biotic: *Erwinia carotovora*, *E. atroseptica*; *Verticillium* wilt. Other: cold induced sweetening	(Carputo *et al.*, 1996, 1997; Frost *et al.*, 2006; Jansky *et al.*, 2011)
	S. brevicaule	Biotic: *Globodera* sp., *G. pallida*, virus	(Ross, 1986; Uhrig and Wenzel, 1981; Santini *et al.*, 2000; Bradshaw and Ramsay, 2005)
	S. candolleanum	Biotic: *Globodera* sp., *G. pallida*, *Erwinia carotovora*, *E. atroseptica*	(Carputo *et al.*, 1996, 1997; Bradshaw and Ramsay, 2005)
	S. vernei	Biotic: virus, pest and nematode	(Ross, 1979, 1986; Bradshaw and Ramsay, 2005)
Secondary	*S. boliviense*	Abiotic: frost	(Tucci *et al.*, 1996; Hawkes *et al.*, 2000)
	S. cajamarquense	Biotic: *Phytophthora infestans*	(Lindqvist-Kreuze *et al.*, 2010)
	S. chacoense	Biotic: virus, pest, *Verticillium* wilt. Other: cold induced sweetening	(Santini *et al.*, 2000; Bradshaw *et al.*, 2006; Frost *et al.*, 2006; Jansky *et al.*, 2011; Narancio *et al.*, 2013)
	S. demissum	Biotic: *Phytophthora infestans*	(Bradshaw *et al.*, 2006; Jo *et al.*, 2011)
	S. kurtzianum	Biotic: *Globodera* sp.	(Bradshaw and Ramsay, 2005)
	S. paucissectum	Biotic: *Phytophthora infestans*	(Villamon *et al.*, 2005)
	S. raphanifolium	Other: cold induced sweetening	(Jansky *et al.*, 2011)
	S. stoloniferum	Biotic: *Phytophthora infestans*, PVY	(Ross, 1979; Bradshaw *et al.*, 2006)
Tertiary	*S. bulbocastanum*	Biotic: *Phytophthora infestans*	(Naess *et al.*, 2000; van der Vossen *et al.*, 2003; Hodgkin and Hajjar, 2008)
	S. commersonii	Biotic: *Ralstonia solanacearum*. Abiotic: frost	(Cardi *et al.*, 1993; Laferriere *et al.*, 1999; González *et al.*, 2013)
	S. palustre	Biotic: PLRV	(Estrada, 1991)
	S. tarnii	Biotic: PVY, *Leptinotarsa decemlineata*, *Phytophthora infestans*	(Thieme *et al.*, 2008)

Uitdewillgen *et al.*, 2013).

Species designations within the section *Petota*, where potato resides, have recently been revised on the basis of new molecular findings in combination with morphological studies (Rodriguez *et al.*, 2009; Spooner, 2009; Ames and Spooner, 2010; Fajardo and Spooner, 2011; Cai *et al.*, 2012; Spooner *et al.*, 2014). The wild related species of potato have been organized into primary, secondary and tertiary gene pools according to the ease of crossability with the cultivated species (Wiersema *et al.*, 2012; Vincent *et al.*, 2013). These wild relatives constitute a morphologically and genetically diverse group of plants distributed from central Chile and Argentina to the southwestern United States. They occupy a variety of habitats within deserts, forests and mountainous regions (Hijmans *et al.*, 2002) (Figure 3.1). Mexico, Bolivia, Argentina, and especially Peru are considered to possess the greatest total diversity of potato wild relatives, although high levels of endemism are reflected in unique species occurring in most of the total 16 countries where these wild relatives grow (Hijmans *et al.*, 2002).

While CWR are likely to play a role in climate change adaptation of novel potato cultivars (Guarino and Lobell, 2011), a number of the wild relatives of cultivated potato are threatened due to habitat destruction and climate change (Jarvis *et al.*, 2008; Vice-ministry for the environment biodiversity and climate change (VMABCC) and Bioversity International, 2009; Cadima *et al.*, 2014). It is therefore becoming more important to address gaps in the *ex situ* conservation of these plants, particularly for species that are currently underrepresented in genebanks and are most impacted in their native habitats.

Gap analysis is a systematic methodology for assessing the comprehensiveness of *ex situ* conservation of plant species, and for assigning taxonomic and geographic priorities for further collecting (Maxted *et al.*, 2008; Ramírez-Villegas *et al.*, 2010). Gap analysis has been applied to the wild relatives of a wide range of crops, including grains, forages and legumes (Ramírez-Villegas *et al.*, 2010; Maxted *et al.*, 2012; Vincent *et al.*, 2012). The analysis can also contribute to the identification of species and habitat priorities for complementary *in situ* conservation.

Here we assessed the current state of *ex situ* conservation of the wild relatives of potato through a gap analysis, in order to identify those species and geographic areas in need of conservation in order to assure their long-term availability for plant breeding efforts.

3.3 Materials and Methods

3.3.1 Wild relative species and geographic area of study

We assessed the closely related wild relatives of potato [i.e., primary and secondary gene pool wild relatives (Harlan and de Wet, 1971)], as well as any distant relatives in the third gene pool that have been published with confirmed or potential uses in crop breeding (Table 3.1). We followed the most recent taxonomic revision of *Solanum* L. section *Petota* (Spooner *et al.*, 2014) (see also Solanaceae Source, `http://solanaceaesource.org/`), henceforth "Solanaceae Source taxonomy". A complementary analysis was also performed following the taxonomy of Ochoa (Ochoa, 1990, 1999, 2003) (henceforth "CIP taxonomy"), in order to provide a gap analysis for the potato wild relative collection conserved as the International Potato Center (CIP), based on its current taxonomic classification (Supplementary Table 3.1). Our study focused on the native distributions of potato wild relatives, which occur in Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Honduras, Mexico, Panama, Paraguay, Peru, Uruguay, USA, and Venezuela (Spooner *et al.*, 2014). Species presence records and germplasm accessions passport data were gathered from online databases (e.g., GBIF -http://www.gbif.org/-, SpeciesLink -http://splink.cria.org.br/-, EURISCO -http://eurisco.ipk-gatersleben.de/-, SINGER, and GRIN -http://www.ars-grin.gov/-), and visits to herbaria (E, K, L, NY and US). Germplasm data were obtained from repositories that provide straightforward access to genetic resources and associated data to the global research community through online information systems. The occurrence data utilized in this analysis is available on (`http://dx.doi.org/10.6084/m9.figshare.1284187`).

3.3.2 Environmental niche modelling

Environmental niche modelling (ENM) techniques were used to estimate the potential geographic distribution of each wild potato species. MaxEnt (Phillips *et al.*, 2006) was selected as the modelling algorithm due to its performance when compared with other modelling approaches, and to its wide use in conservation analyses (Elith *et al.*, 2006; Simon *et al.*, 2010, 2011). Ten thousand random points were used as background records across Central and South America, the native range of the wild relatives. A five-fold cross-validation option (k=5) was implemented to maximize the use of small sets of georeferenced records in the modelling, producing five replicates per species, subsequently summarized into a single ensemble model by estimating the mean values across the replicates. The models were restricted to their known native countries per species as reported in the literature (Spooner *et al.*, 2014), and further refined using a species-specific threshold corresponding to the shortest distance to the upper left corner of the Receiver Operating Characteristic (ROC) curve (Liu *et al.*, 2005). For environmental drivers, we used 19 bioclimatic variables (Supplementary Table 3.2) derived from the WorldClim database (Hijmans *et al.*, 2005) at a resolution of 2.5 arc-minutes (approx. 5 km at the equator).

The performance of each ENM was assessed to determine its suitability for use in the gap analysis. Three parameters were checked: (i) the 5-fold average Area Under the Test ROC Curve (ATAUC), (ii) the standard deviation of the ATAUC for the 5 different folds, and (iii) the proportion of potential distribution where the standard deviation is greater than 0.15 (ASD15). A suitable model had to meet these conditions: ATAUC > 0.7, STAUC < 0.15 and ASD15 < 10% (Ramírez-Villegas *et al.*, 2010). In those cases where a suitable niche model was not produced (either due to lack of data or low performance of the ensemble model), a convex hull (polygon surrounding the outermost georeferenced points) was prepared.

3.3.3 Gap analysis

We used a gap analysis methodology (Maxted *et al.*, 2008; Ramírez-Villegas *et al.*, 2010) including three metrics to determine the urgency of collecting wild relatives for conservation *ex situ*. A Sampling Representativeness Score (SRS) compared the number of germplasm accessions to the total number of samples (germplasm plus species presence records, with or without geographic coordinates), giving a general overview of the sufficiency of accessions per species. A Geographic Representativeness Score (GRS) compared the ENMs of the species to the geographic distribution of existing germplasm accession collecting sites, estimated by creating circular buffers of 50 km (CA50) around each site where the accession was collected (Hijmans and Spooner, 2001), in order to assess the geographic coverage of germplasm collections. An Ecosystem Representativeness Score (ERS) assessed the number of ecosystems currently represented in *ex situ* collections (CA50 of germplasm collections), in comparison to the total number of ecosystems distributed within the ENMs of species. The three gap analysis metrics were given equal weight and an average was calculated to obtain a Final Priority Score (FPS). Four categories were employed to assign priority for further collecting for *ex situ* conservation: high priority species (HPS) when $\text{FPS} \leq 3$, or when ten or less accessions were recorded in germplasm collections; medium-priority species (MPS) when $3 < \text{FPS} \geq 5$; low priority species (LPS) when $5 < \text{FPS} \geq 7.5$; and "no further collecting of germplasm required" (NFCR) when $7.5 < \text{FPS} \geq 10$.

The gap analysis was performed using R v2.15.1 (R Core Team, 2014), and the packages maptools (Bivand and Lewin-Koh, 2014), rgdal (Bivand *et al.*, 2014), SDMTools (VanDerWal *et al.*, 2014), raster (Hijmans, 2014), sp (Pebesma and Bivand, 2005; Bivand *et al.*, 2013), dismo (Hijmans *et al.*, 2014) and ggplot2 (Wickham, 2009).

3.3.4 Identification of geographic areas of priority for further collecting

Maps highlighting areas identified as priorities for further collecting (collecting gaps) were prepared for each species by subtracting the existing germplasm CA50 buffers from the ENMs. For those species where a niche model was not produced, CA50 buffers were prepared around all presence records, with germplasm CA50 buffers subtracted from these representations of the distribution of species. Collecting gap maps for all high priority species were analyzed using the "Zonal Statistics" tool in ArcMap (ESRI, 2011) to produce a count of species in need of further collecting per country.

3.4 Results

3.4.1 Wild relative species and geographic area of study

Seventy-three species were included in the analysis as relatively close relatives of potato (i.e., members of the primary and secondary gene pools (Harlan and de Wet, 1971) or due to published actual or potential use in breeding efforts. These included seven species from the primary gene pool of potato, 63 from the secondary gene pool, and three tertiary gene pool species with reported use in crop improvement (Table 3.2). Almost half of the species analyzed are diploids with an endosperm balance number of 2 (2 EBN), followed by tetraploids (2 EBN and 4 EBN) and hexaploids (4 EBN) (Spooner and Hijmans, 2001). For the complementary gap analysis, following the CIP taxonomy, a total of 187 putative species were analyzed, equivalent to the 73 Solanaceae Source taxonomy species (Spooner *et al.*, 2014) (Supplementary Table 3.1). A total of 49,164 records for the 73 potato wild relatives were gathered (75.76% with coordinates), with 11,100 germplasm accessions and 37,251 presence records, including herbarium references, inactive germplasm accessions, and field sighting recordings (Fig. 3.2A).

Figure 3.1: Flowers, plants and habitats of A) *Solanum acaule*, B) *S. candolleanum*, C) *S. laxissimum*, D) *S. rhomboideilanceolatum*, E) *S. simplicissimum* and F) *S. wittmackii*. Photographs by S. de Haan. The author of the photographs has given written consent to publish them.

Table 3.2: List of 73 species analyzed and their corresponding prioritization category, gene pool, ploidy level, native areas and count of . data retrieved for this study. SRS: Sampling Representativeness Score, GRS: Geographical Representativeness Score, ERS: Environmental Representativeness Score, FPCAT: Final priority category. HPS = high priority species, MPS = medium priority species, LPS = low priority species, and NFCR = no further collecting required. GP: gene pool, 1: Primary, 2: Secondary, and 3: Tertiary. AR: Argentina, BO: Bolivia, BR: Brazil, CL: Chile, CO: Colombia, CR: Costa Rica, EC: Ecuador, GT: Guatemala, HN: Honduras, MX: Mexico, PA: Panama, PE: Peru, PY: Paraguay, UY: Uruguay, US: United States of America, and VE: Venezuela

Taxon (Solanaceae Source)	**Countries**	**Ploidy (Hijmans *et al.*, 2007) and (EBN) (Spooner and Hijmans, 2001)**	**GP**	**No. of reference samples (georeferenced)**	**No. of germplasm accessions (georeferenced)**	**SRS**	**GRS**	**ERS**	**FPS**	**FPCAT**
S. acaule Bitter	AR; BO; PE; CL	4x (2EBN), 6x	1	3058 (864)	1762 (521)	3.66	10	10	7.89	NFCR
S. acroglossum Juz.	PE	2x (2EBN)	2	92 (23)	4 (4)	0.42	0.61	3	0	HPS
S. acroscopicum Ochoa	PE	2x	2	93 (38)	11 (7)	1.06	0.9	6.36	2.77	HPS
S. agrimonifolium Rydberg	GT; HN; MX	4x (2EBN)	2	345 (118)	40 (14)	1.04	6.48	4.21	3.91	MPS
S. albicans (Ochoa) Ochoa	EC; PE	6x (4EBN)	2	288 (73)	132 (40)	3.14	5.2	10	6.11	LPS
S. albornozii Correll	EC	2x (2EBN)	2	25 (7)	13 (8)	3.42	5.06	7.5	5.33	LPS
S. andreanum Baker	CO; EC	2x (2EBN); 4x (4EBN)	2	448 (234)	111 (71)	1.99	5.06	6.47	4.51	MPS
S. ayacuchense Ochoa	PE	2x (2EBN)	2	10 (7)	0 (0)	0	0	0	0	HPS
S. berthaultii J. G. Hawkes	AR; BO	2x (2EBN), 3x	1	836 (292)	323 (116)	2.79	7.68	10	6.82	LPS
S. boliviense M. F. Dunal in DC.	BO; PE; AR	2x (2EBN)	2	1724 (657)	388 (185)	1.84	8	10	6.61	LPS
S. bombycinum C. M. Ochoa	BO	4x	2	8 (6)	1 (1)	1.11	1.62	5	0	HPS
S. brevicaule Bitter	AR; BO; PE	2x (2EBN); 4x (4EBN); 6x (4EBN)	1	4428 (1477)	1159 (457)	2.07	10	10	7.36	LPS
S. buesii Vargas	PE	2x (2EBN)	2	63 (32)	6 (4)	0.87	0.24	2.73	0	HPS
S. bulbocastanum Dunal in Poiret	GT; MX	2x (1EBN), 3x	3	970 (399)	175 (47)	1.53	6.2	10	5.91	LPS
S. burkartii Ochoa	PE	2x	2	88 (18)	7 (5)	0.74	6.09	8.33	0	HPS
S. cajamarquense Ochoa	PE	2x (1EBN)	2	223 (39)	16 (8)	0.67	1.06	6	2.58	HPS
S. candolleanum Berthault	PE; BO	2x (2EBN), 3x	1	2910 (1245)	739 (349)	2.03	10	9.17	7.06	LPS
S. cantense Ochoa	PE	2x (2EBN)	2	155 (68)	3 (3)	0.19	0.93	3.75	0	HPS
S. chacoense Bitter	AR; BO; PY; PE; UY; BR	2x (2EBN), 3x	2	2527 (1004)	710 (119)	2.19	1.94	5.52	3.22	MPS
S. chilliasense Ochoa	EC	2x (2EBN)	2	15 (7)	5 (4)	2.5	10	10	0	HPS
S. chiquidenum Ochoa	PE	2x (2EBN)	2	360 (148)	17 (11)	0.45	3.27	7	3.57	MPS
S. chomatophilum Bitter	PE; EC	2x (2EBN)	2	967 (378)	124 (55)	1.14	6.54	8.33	5.34	LPS
S. clarum D. S. Correll	GT; MX	2x	2	244 (92)	6 (4)	0.24	3.69	2.78	0	HPS
S. colombianum Dunal	CO; EC; PA; VE	4x (2EBN)	2	1116 (444)	214 (105)	1.61	6.47	9.14	5.74	LPS
S. commersonii M. F. Dunal	AR; BR; UY	2x (1EBN), 3x	3	692 (272)	112 (30)	1.39	2.14	5.83	3.12	MPS
S. contumazaense Ochoa	PE	2x (2EBN)	2	21 (13)	2 (2)	0.87	5.26	6.67	0	HPS
S. demissum Lindley	GT; MX	6x (4EBN)	2	1669 (513)	613 (85)	2.69	8.3	10	6.99	LPS

Table 3.2 – continued from previous page

Taxon (Solanaceae Source)	**Countries**	**Ploidy (Hijmans *et al.*, 2007) and (EBN)(Spooner and Hijmans, 2001)**	**GP**	**No. of reference samples (georeferenced)**	**No. of germplasm accessions (georeferenced)**	**SRS**	**GRS**	**ERS**	**FPS**	**FPCAT**
S. flahaultii Bitter	CO	4x	2	99 (37)	39 (10)	2.83	2.66	3.75	3.08	MPS
S. gandarillasii H. M. Cárdenas	BO	2x (2EBN)	2	48 (28)	21 (7)	3.04	3.72	7.14	4.64	MPS
S. garcia-barrigae Ochoa	CO	4x	2	21 (10)	3 (2)	1.25	0.52	1.9	0	HPS
S. gracilifrons Bitter	PE	2x	2	19 (8)	1 (1)	0.5	1.47	3.75	0	HPS
S. guerreroense D. S. Correll	MX	6x (4EBN)	2	4 (2)	20 (2)	8.33	10	10	9.44	NFCR
S. hastiforme Correll	PE	2x (2EBN)	2	49 (32)	2 (2)	0.39	0.38	4	0	HPS
S. hintonii D. S. Correll	MX	2x	2	39 (18)	0 (0)	0	0	0	0	HPS
S. hjertingii J. G. Hawkes	MX	4x (2EBN)	2	155 (62)	54 (10)	2.58	1.93	4	2.84	HPS
S. hougasii D. S. Correll	MX	6x (4EBN)	2	186 (79)	39 (10)	1.73	2.12	3.68	2.51	HPS
S. huancabambense Ochoa	PE	2x (2EBN)	2	111 (28)	29 (10)	2.07	2.07	5.56	3.23	MPS
S. incasicum Ochoa	PE	2x (2EBN)	2	9 (5)	2 (2)	1.82	10	5	0	HPS
S. infundibuliforme R. A. Philippi	AR; BO	2x (2EBN)	1	836 (277)	234 (116)	2.19	4.71	7.78	4.89	MPS
S. iopetalum (Bitter) J. G. Hawkes	MX	6x (4EBN)	2	626 (313)	93 (51)	1.29	5.23	7.5	4.67	MPS
S. kurtzianum Bitter & L. Wittmack	AR	2x (2EBN)	2	764 (253)	276 (32)	2.65	4.02	8.75	5.14	LPS
S. laxissimum Bitter	PE	2x (2EBN)	2	139 (91)	19 (10)	1.2	1.73	5	2.64	HPS
S. lesteri J. G. Hawkes & Hjerting	MX	2x	2	23 (12)	12 (4)	3.43	4.22	4.44	4.03	MPS
S. limbaniense Ochoa	PE	2x (2EBN)	2	56 (28)	12 (7)	1.76	1.18	5	2.65	HPS
S. lobbianum Bitter	CO	4x (2EBN)	2	1 (1)	4 (1)	8	NA	NA	0	HPS
S. longiconicum Bitter	CR; PA	4x	2	546 (198)	25 (12)	0.44	10	10	6.81	LPS
S. maglia D. F. L. von Schlechtendal	CL; AR	2x, 3x	2	190 (51)	15 (4)	0.73	0.14	1.33	0.74	HPS
S. medians Bitter	PE; CL	2x (2EBN), 3x	2	849 (305)	98 (35)	1.03	4.32	4.44	3.27	MPS
S. microdontum Bitter	AR; BO	2x (2EBN), 3x	2	1178 (349)	422 (94)	2.64	6.25	9.09	5.99	LPS
S. morelliforme Bitter & Muench	GT; MX; HN	2x	2	364 (140)	45 (18)	1.1	4.74	6.55	4.13	MPS
S. multiinterruptum Bitter	PE	2x (2EBN), 3x	2	496 (204)	95 (45)	1.61	7.33	8.75	5.9	LPS
S. neocardenasii J. G. Hawkes & J. P. Hjerting	BO	2x	2	25 (17)	17 (5)	4.05	0.56	3.64	2.75	HPS
S. neorossii J. G. Hawkes & J. P. Hjerting	AR	2x	2	76 (35)	45 (14)	3.72	4.17	10	5.96	LPS
S. neovavilovii Ochoa	BO	2x (2EBN)	2	26 (13)	0 (0)	0	0	0	0	HPS
S. nubicola Ochoa	PE	4x (2EBN)	2	36 (20)	2 (2)	0.53	0.7	5.45	0	HPS
S. okadae J. G. Hawkes & J. P. Hjerting	BO	2x	1	139 (55)	75 (19)	3.5	1.08	7.14	3.91	MPS
S. olmosense Ochoa	EC; PE	2x (2EBN)	2	26 (15)	0 (0)	0	0	0	0	HPS
S. oxycarpum Schiede in D. F. L. von Schlechtendal	MX	4x (2EBN)	2	203 (77)	58 (20)	2.22	2.45	7.93	4.2	MPS
S. paucissectum Ochoa	PE	2x (2EBN)	2	182 (20)	20 (10)	0.99	10	10	7	LPS
S. pillahuatense Vargas	PE	2x (2EBN)	2	15 (11)	1 (1)	0.63	10	10	0	HPS
S. piurae Bitter	PE	2x (2EBN)	2	226 (38)	17 (7)	0.7	0.47	3	1.39	HPS
S. polyadenium Greenman	MX	2x	2	286 (97)	99 (14)	2.57	3.52	8.13	4.74	MPS
S. raphanifolium Cárdenas & Hawkes	PE	2x (2EBN)	2	597 (206)	220 (69)	2.69	6.52	8.57	5.93	LPS

Table 3.2 – continued from previous page

Taxon (Solanaceae Source)	**Countries**	**Ploidy (Hijmans *et al.*, 2007) and (EBN)(Spooner and Hijmans, 2001)**	**GP**	**No. of reference samples (georeferenced)**	**No. of germplasm accessions (georeferenced)**	**SRS**	**GRS**	**ERS**	**FPS**	**FPCAT**
S. rhomboideilanceolatum Ochoa	PE	2x (2EBN)	2	99 (46)	7 (3)	0.66	1.04	6.67	0	HPS
S. salasianum Ochoa	PE	2x	2	13 (7)	0 (0)	0	0	0	0	HPS
S. schenckii Bitter	MX	6x (4EBN)	2	105 (37)	49 (13)	3.18	2.45	6.8	4.14	MPS
S. sogarandinum Ochoa	PE	2x (2EBN), 3x	2	157 (81)	27 (13)	1.47	3.22	6.67	3.79	MPS
S. stoloniferum D. F. L. von Schlechtendal	MX; US	4x (2EBN)	2	3807 (1464)	1582 (314)	2.94	10	10	7.65	NFCR
S. tarnii J. G. Hawkes & Hjerting	MX	2x	3	68 (31)	45 (10)	3.98	2.58	4.62	3.73	MPS
S. venturii J. G. Hawkes & J. P. Hjerting	AR	2x (2EBN)	2	165 (62)	39 (6)	1.91	0.47	4.44	2.28	HPS
S. vernei Bitter & L. Wittmack	AR	2x (2EBN)	1	429 (122)	261 (47)	3.78	2.46	8.89	5.04	LPS
S. verrucosum D. F. L. von Schlechtendal	MX	2x (2EBN), 3x, 4x	2	968 (378)	222 (36)	1.87	6.56	5.91	4.78	MPS
S. violaceimarmoratum Bitter	BO; PE	2x (2EBN)	2	234 (104)	61 (16)	2.07	0.98	2.86	1.97	HPS

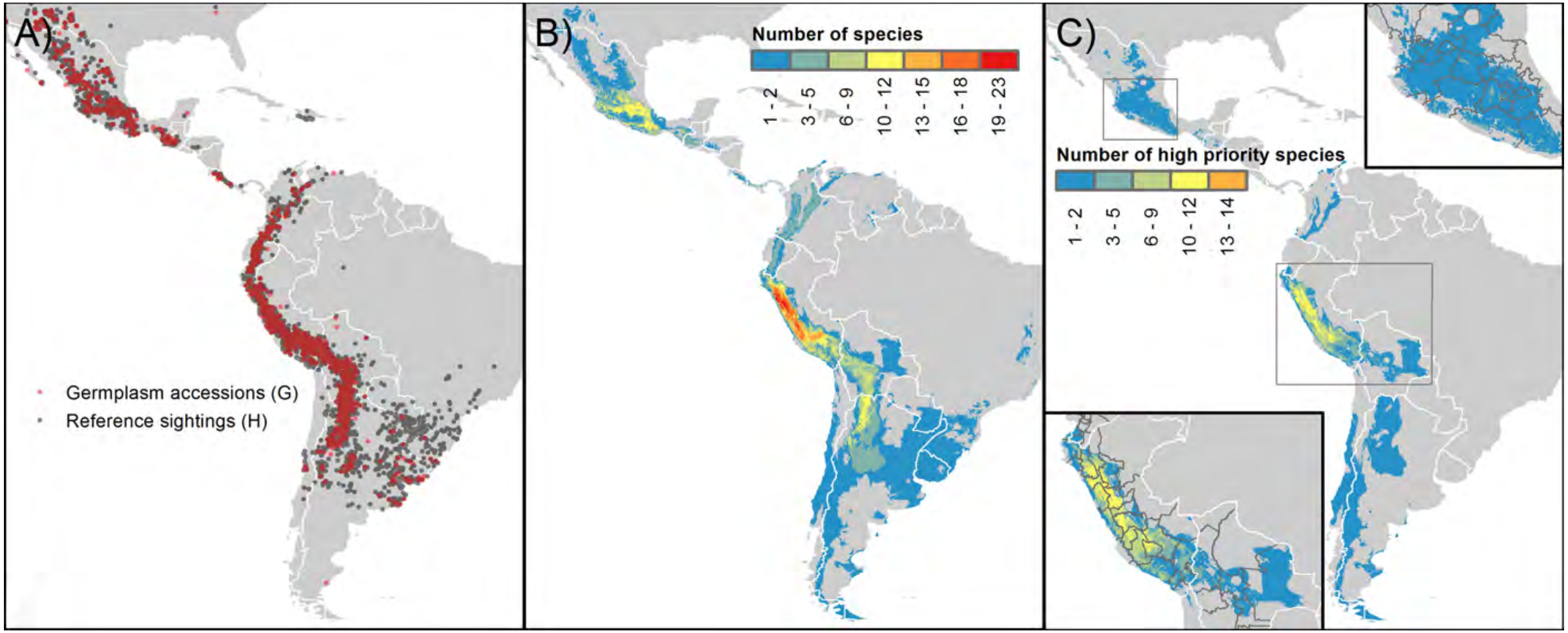


Figure 3.2: Distribution of the wild relatives of potato: A) Distribution of germplasm and herbarium records included in the analysis. Red dots represent germplasm accessions (G) and dark gray dots herbarium/presence records (H). B) Species richness based upon environmental niche models, and C) Potential hotspots for further collecting of high priority species (HPS).

3.4.2 Environmental niche modelling

The environmental niche models of 75 species (89%) met the parameters used to consider an ENM suitable for use in the gap analysis. For the remaining eight species (*S. chilliasense*, *S. guerreroense*, *S. incasicum*, *S. lobbianum*, *S. neovavilovii*, *S. olmosense*, *S. paucissectum*, and *S. pillahuatense*), convex hulls were prepared and used in the gap analysis, as the ENM replicates produced were highly variable and did not comply with the ASD15 condition. Potato crop wild relative species richness was found to be highest in Peru, followed by Mexico and Argentina (Fig. 3.2B, Supplementary File 3.1).

Occurrence data, ENMs and the collecting priorities maps for the species analyzed, following the Solanaceae Source taxonomy, are available in an interactive format at `http://www.cwrdiversity.org/distribution-map/`

3.4.3 Gap analysis

The gap analysis for the 73 species resulted in the assignment of 32 HPS, 20 MPS, 18 LPS and 3 NFCR (Table 3.2). There are no germplasm accessions currently available for *S. ayacuchense*, *S. neovavilovii*, *S. olmosense* and *S. salasianum*, and these species therefore represent the greatest urgency for further collecting. All HPS belong to the secondary gene pool (Fig. 3.3).

Solanum neocardenasii and *S. lobbianum* possessed a single dominant factor contributing to their priority category assignment for further collecting. All other species possessed two (40.6% of the species), three (28.1%) or four (28.1%) factors contributing importantly to their FPS status (Supplementary Table 3.3). Ninety-four percent of the species classified as HPS had a low SRS (SRS equal or less than 3) [median (mean) = 0.73 (1.22)] (Fig. 3.4A, Supplementary Fig. 3.1). Likewise, 78.1% of HPS exhibited a low GRS [0.930 (2.07)] (Supplementary Fig. 3.1), with five species well represented (*S. candolleanum*, *S. brevicaule*, *S. stoloniferum* and *S. acaule*), as shown in Fig 3.4B, where the dashed line is the complete representativeness line, and the continuous line is the average representativeness line, the former showing an ideal scenario where the potential geographic extension of the gene pool is completely represented

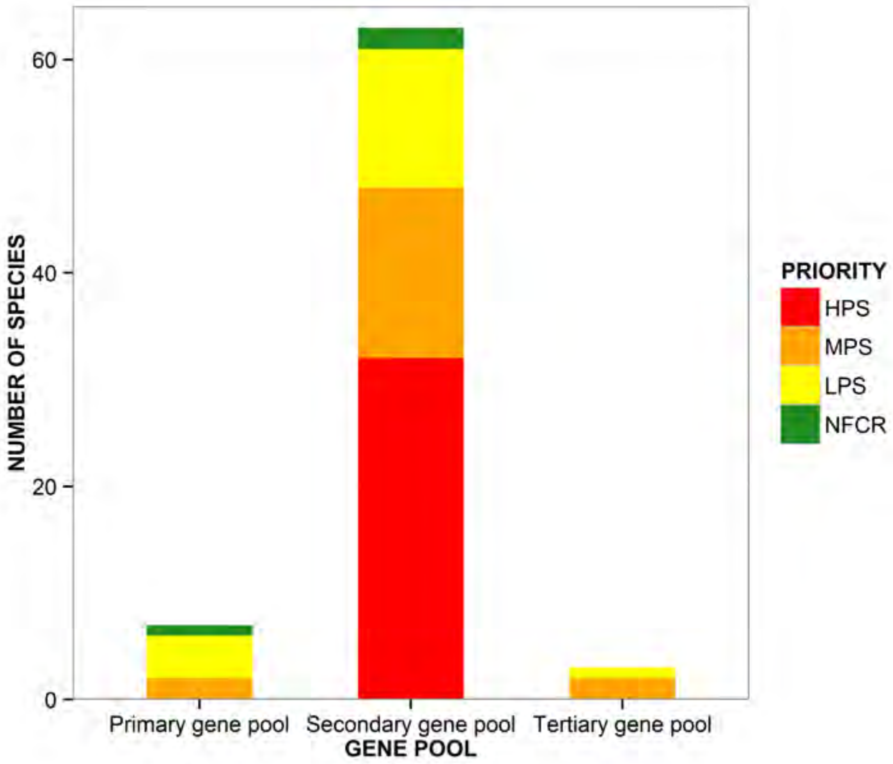


Figure 3.3: Priorities for further collecting by potato wild relative gene pool, categorized into high priority species (HPS), medium priority species (LPS), low priority species (LPS), and "no further collecting required" (NFCR).

at genebank collections and the latter showing the extent of representativeness compared to the potential extent of the gene pool. On the other hand, the ERS contributed less to the FPS of high priority species, with less than half (37.5%) of the HPS exhibiting an ERS ≤ 3 [median value 3.75 (4.01)] (Fig. 3.4C, Supplementary Fig. 3.1). A total of 65.6% of the species ranked as high priority had less than ten active accessions and consequently very limited representativeness in terms of absolute numbers of accessions available in germplasm collections.

A total of 31 HPS were mapped together for targeting of geographic hotspots for further collecting (Fig. 3.2C, Supplementary File 3.2). Peru contained the highest count of HPS for further collecting (21 species), followed by Mexico (4); Bolivia (3); Colombia (2), Ecuador (2) and Argentina, Chile and Guatemala (each with 1 species) (Fig. 3.2C). Twenty-eight species (out of 32) were found to be endemic to a single country (Fig. 3.5). The greatest concentrations of species requiring further collecting were predicted to occur in the Peruvian Departments of

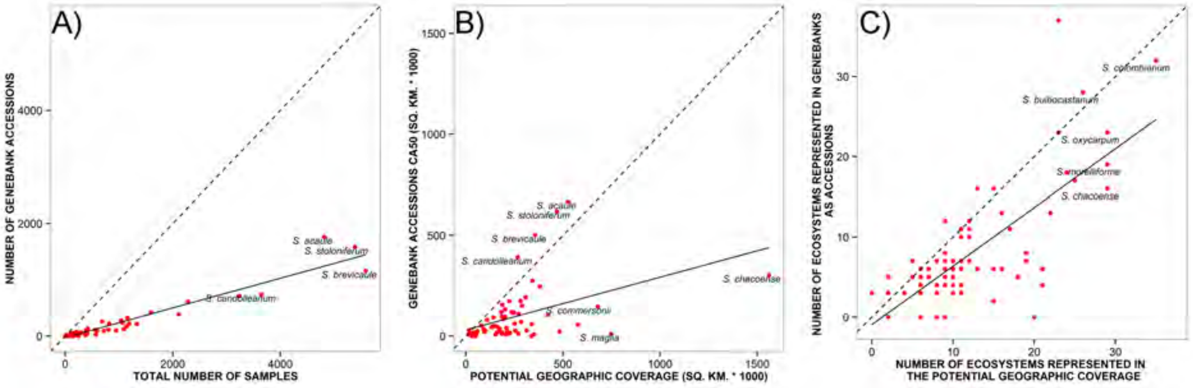


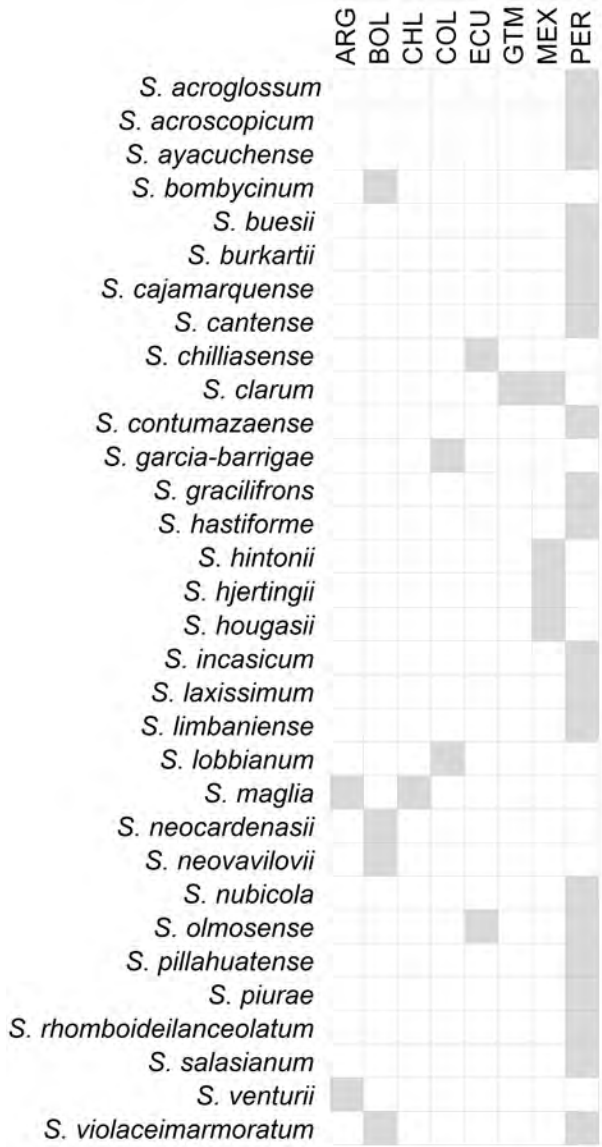
Figure 3.4: Gap analysis metrics obtained for all the crop wild relatives analyzed following the Solanaceae Source taxonomy. A) Sampling Representativeness Score (SRS), B) Geographic Representativeness Score (GRS), C) Ecosystem Representativeness Score (ERS).

Cajamarca, La Libertad, Ancash and Huánuco. Supplementary Table 3.4 provides an overview of sites recommended for further collecting of high priority species based on their presence points.

A total of 18 species were assessed as MPS for further collecting, and are distributed in: Argentina (1 species), Bolivia (2), Colombia (1), Ecuador (2), Guatemala (2), Mexico (8), Peru (5), Honduras (2), Paraguay (1), Uruguay (1) and Brazil (1) (Fig. 3.6).

The restricted range and endemic nature of many of the insufficiently collected taxa implies that targeted collecting trips to specific regions outside the gap richness areas are needed in order to form comprehensive germplasm collections for potato wild relatives. Some of the HPS species are known to occur in threatened habits, requiring urgent attention (e.g., *S. rhomboideilanceolatum* (Fig. 3.1D) and *S. piurae*). Other species, such as *S. laxissimum* (Fig. 3.1C) and *S. neovavilovii*, occur in relatively intact natural areas or within the boundaries of national parks and can thus be expected to be more secure. Active monitoring of these species in the wild can provide greater assurance of continued conservation in these areas.

Following the CIP taxonomy system, 65.1% of the 172 species analyzed were ranked as high priorities for further collecting (HPS), 19.8% were identified as medium priorities (MPS), 9.9% as low priorities and 5.2% as "no further collecting required" (NFCR) (Supplementary Figure 3.2).



	ARG	BOL	CHL	COL	ECU	GTM	MEX	PER
S. acroglossum								▨
S. acroscopicum								▨
S. ayacuchense								▨
S. bombycinum		▨						
S. buesii								▨
S. burkartii								▨
S. cajamarquense								▨
S. cantense								▨
S. chilliasense					▨			
S. clarum						▨	▨	
S. contumazaense								▨
S. garcia-barrigae				▨				
S. gracilifrons								▨
S. hastiforme								▨
S. hintonii							▨	
S. hjertingii							▨	
S. hougasii							▨	
S. incasicum								▨
S. laxissimum								▨
S. limbaniense								▨
S. lobbianum				▨				
S. maglia	▨		▨					
S. neocardenasii		▨						
S. neovavilovii		▨						
S. nubicola								▨
S. olmosense					▨			▨
S. pillahuatense								▨
S. piurae								▨
S. rhomboideilanceolatum								▨
S. salasianum								▨
S. venturii	▨							
S. violaceimarmoratum		▨						▨

Figure 3.5: Countries identified for potential further collecting per high priority crop wild relative species. ARG: Argentina, BOL: Bolivia, CHL: Chile, COL: Colombia, ECU: Ecuador, GTM: Guatemala, MEX: Mexico, PER: Peru.

3.5 Discussion

With 32 species classified as high priority and another 20 as medium priority for collecting, it is evident that further conservation action is needed to safeguard the wild genetic resources of this globally important crop. We propose three levels of priority for further collecting: first for the four HPS species that are completely lacking from internationally available genebank collections (*S. ayacuchense*, *S. neovavilovii*, *S. olmosense* and *S. salasianum*); second for the

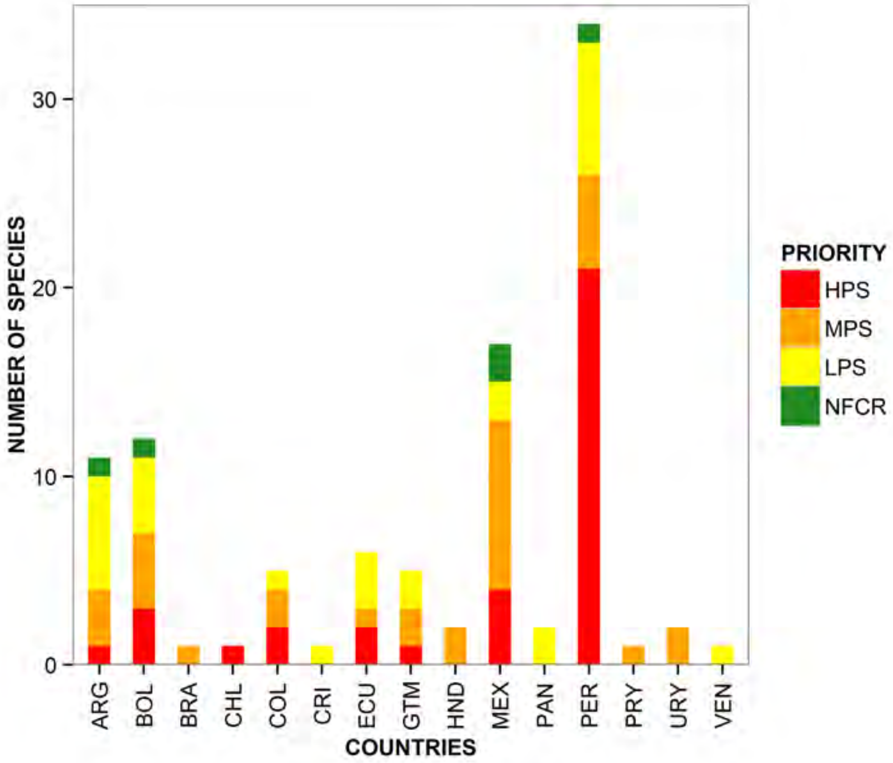


Figure 3.6: Number of CWR species prioritized for further collecting per country. HPS = high priority species, MPS = medium priority species, LPS = low priority species, and NFCR = "no further collecting required" (NFCR). ARG: Argentina, BOL: Bolivia, BRA: Brazil, CHL: Chile, COL: Colombia, CRI: Costa Rica, ECU: Ecuador, GTM: Guatemala, HND: Honduras, MEX: Mexico, PAN: Panama, PER: Peru, PRY: Paraguay, URY: Uruguay and VEN: Venezuela.

other 28 HPS species occurring in a total of eight countries; and third for the MPS.

In addition to gap filling for *ex situ* collections, the results can help establish priorities for the establishment of genetic reserves for the *in situ* conservation of potato wild relatives. Such reserves may most effectively be established at sites where several HPS and/or MPS overlap, especially if coinciding with existing protected areas. Habitats undergoing significant disturbance may also represent high priorities for consideration for *in situ* conservation efforts.

Some of the HPS display very restricted distributions and are considered to be threatened *in situ*. The limited habitat of *S. rhomboideilanceolatum* in Peru is increasingly exposed to road building and overgrazing by livestock (field observation by the authors, 2013). Yet other HPS with restricted distributions, such as *S. bombycinum* in Bolivia, are reported to grow in habitats that are not presently highly exposed to threats (Cadima *et al.*, 2014), while additional species with relatively extensive ranges such as *S. laxissimum* in Peru show considerable spatial overlap with protected areas. Factors such as threats to the *in situ* conservation of wild populations, overlap with protected areas, and degree of endemism can further refine collecting priorities. Monitoring the population dynamics, ecology and genetics of selected species to corroborate the effect of climate change and other threats to wild relatives also represent useful contributions to conservation planning (Iriondo *et al.*, 2012). Such studies can help to ground-truth climate change forecasts and to enhance the understanding of the adaptive capacity of wild relatives.

Many of the taxa classified as generally well conserved (LPS and NFCR) are those that are widely used in breeding programs, such as *S. bulbocastanum* and *S. stoloniferum*. This is a logical consequence of demand from such programs. It is anticipated that demand for as yet underutilized species will increase as potato breeding efforts expand the use of wide diversity in order to confront emerging biotic and abiotic stresses.

Our results assign a relatively large number of species from Peru to the category of high priority for further collecting. This may seem surprising given the long history of collecting missions in the center of species diversity. Sampling biases relative to road systems, time limitations of collecting missions and the tendency of collectors to sample in areas of previous expeditions have been reported (Hijmans *et al.*, 2000, 2002). The restricted distribution, high levels of endemism, and difficult access to some of the areas where HPS potato wild relatives occur provide further insight into the low level of representation of a number of these species in genebanks. New roads in Peru in previously isolated and remote habitats will soon make these populations increasingly accessible for collecting but at the same time more vulnerable to habitat destruction.

Long-term conservation of the genetic diversity of wild relatives of potato will also require further research in population genetics and reproductive biology of the species (Ovchinnikova *et al.*, 2011). Gap filling of the taxa identified here as critically under-represented in germplasm collections will provide an important step in making germplasm available for such analyses. Future studies should incorporate morphological and molecular analyses in order to elucidate the diversity and genetic distances within and between populations of wild relatives as well as between genebank collections and *in situ* reserves (del Rio *et al.*, 1997a,b; Camadro, 2014). Genetic variability encountered within natural populations has typically been poorly studied and is generally not taken into account before planning collecting expeditions for wild relatives (Camadro, 2012). Further taxonomic research may also be useful. The complementary gap analysis following the CIP taxonomy displayed differences in resulting priorities for further collecting (Supplementary Figure 3.2), and may reveal potentially useful infraspecific variation for further exploration, as some of the species in CIP taxonomy may represent unique subpopulations within the Solanaceae Source taxonomy.

The collecting priorities identified here, combined with further emphasis on improving *ex situ* conservation technologies and associated data management, performing genotypic and phenotypic characterization of wild relative diversity, monitoring wild populations *in situ*, and making conserved wild relatives and their associated data accessible to the global research community, represent key steps in ensuring the long-term availability of the wild genetic resources of this critically important crop.

CHAPTER 4

CROP WILD RELATIVES OF THE BRINJAL EGGPLANT (*SOLANUM MELONGENA*: SOLANACEAE): POORLY REPRESENTED IN GENEBANKS AND MANY SPECIES AT RISK OF EXTINCTION

Article accepted for publication by the American Journal of Botany.

Mindy M. Syfert, **Nora P. Castañeda-Álvarez**, Colin K. Khoury, Tiina Särkinen, Chrystian C. Sosa, Harold A. Achicanoy, Vivian Bernau, Jaime Prohens, Marie-Christine Daunay and Sandra Knapp.

MMS, SK and NPC-A conceived and designed the study. MMS, TS and NPC-A collated and prepared the data. MMS, NPC-A, CCS and HAA performed the analysis. MMS and NPC-A interpreted the results. MMS, SK and NPC-A* wrote the manuscript. MMS, NPC-A, CKK, TS, CCS, HAA, VB, JP, MCD and SK critically reviewed and edited the manuscript.

*This chapter was written in collaboration with Dr. Mindy Syfert and Dr. Sandra Knapp. Full details of our contributions are given in the Declaration of this thesis.

4.1 Abstract

- *Premise of the study.* Crop wild relatives (CWR) provide important traits for plant breeding, including pest, pathogen and abiotic stress resistance. Their conservation and future availability are essential for food security. Despite this, the world's genebanks are currently thought to conserve only a fraction of the total diversity of CWR.

- *Methods.* We define the eggplant gene pool using the results of recent taxonomic and phylogenetic studies. We identify the gaps in germplasm accessions for eggplant (*Solanum melongena* L.) CWR by comparing georeferenced herbarium records and germplasm accessions using a gap analysis methodology implementing species distribution models (SDM). Preliminary conservation assessments using IUCN criteria were done for all species and were combined with the gap analysis to pinpoint where under-collected and threatened CWR species coincide with high human disturbance and occur outside of protected areas.

- *Key results.* We show that many eggplant CWR are poorly represented in genebanks compared to their native ranges. Priority areas for future collection are concentrated in Africa, especially along the Kenya-Tanzania border. Fourteen species of eggplant CWR are assessed as threatened or near-threatened; these are also concentrated in eastern Africa.

- *Conclusions.* The knowledge base upon which conservation of wild relative germplasm depends must take into account both taxonomic and phylogenetic advances. Beyond traditional research focus on close relatives of crops, we emphasize the benefits of defining a broad CWR gene pool, and the importance of assessing threats to wild species when targeting localities for future collection of CWR to improve crop breeding in the face of environmental change.

- *Key words.* Agriculture; crop gene pools; gap analysis; genebanks; eggplant; food security; Red List assessments; species distribution modelling; Solanaceae; *Solanum*

4.2 Introduction

Domestication and subsequent improvement has generally led to narrowing of the genetic diversity of our commonly cultivated crop varieties (Tanksley and McCouch, 1997). Limited genetic diversity also limits the potential of crop breeding efforts, especially in the face of environmental change. This situation has re-invigorated research into the use of crop wild relatives (CWR), the wild cousins of domesticated crop species, in breeding efforts (Tanksley and McCouch, 1997; Hodgkin and Hajjar, 2008; Dempewolf *et al.*, 2013). Crop wild relatives share a relatively recent common ancestry with domesticated species and due to that close relationship, are reservoirs of genetic traits that can be useful in crop improvement (Tanksley and McCouch, 1997; Guarino and Lobell, 2011). Interest in CWR has its roots in Nikolai Vavilov's centers of origin of cultivated plants (as compiled in Vavilov, 1992) concepts where regions identified as rich in wild species related to crops were suggested to be the centers of domestication and origin of these crops, and where significant genetic diversity was thought to reside, both in terms of related wild species and in the crop itself. Despite this long history of interest and the recognized value of wild and weedy crop relatives, they remain largely understudied and their natural populations are increasingly at risk due to habitat loss and conversion, climate change, and invasive species (e.g., Ford-Lloyd *et al.*, 2011).

In the past decade, the International Treaty on Plant Genetic Resources for Food and Agriculture (FAO, 2009) has provided momentum for efforts to better conserve and share such wild genetic resources. More recently, international instruments such as the United Nations' Sustainable Development Goals, the updated Global Strategy for Plant Conservation (CBD, 2010) and the Aichi Biodiversity Targets (Secretariat of the CBD, 2010) have specifically targeted the comprehensive conservation of wild plant genetic resources.

The conservation of CWR is a matter of urgency because many of these species are considered to be threatened in their natural habitats due to the conversion of land to industrial agricultural production, urbanization, mining, the spread of invasive species, pollution, and climate change (Jarvis *et al.*, 2008; Brummitt *et al.*, 2015a). A number of initiatives have been enacted to conserve such genetic resources in the wild, but these efforts have treated a limited number of species and geographic regions (Meilleur and Hodgkin, 2004), and much greater efforts are needed to adequately conserve wild genetic resources as a whole (Maxted and Kell, 2009; Iriondo *et al.*, 2012).

Genebanks, also known as seedbanks, have been an important component in the conservation and availability of wild species useful for plant breeding (Tanksley and McCouch, 1997), enabling their maintenance in controlled environments and facilitating their availability to crop breeders and other researchers. Assessing the level of representation of wild species in these genebanks is thus important for evaluating the state of conservation of these species, and for guiding future efforts to fill important gaps in these collections (Maxted *et al.*, 2008; Vimal *et al.*, 2010). Gap analysis methodologies perform such assessments by identifying those wild taxa, geographic locations, and particular traits or adaptations that are un- or under-conserved in genebanks (Maxted *et al.*, 2008) based upon available information resources (e.g., occurrence records from herbaria and genebanks) and using geographic and species distribution modelling methods (Ramírez-Villegas *et al.*, 2010; Parra-Quijano *et al.*, 2011; Castañeda-Álvarez *et al.*, 2016).

Efforts to protect wild plant genetic resources have been focused on major cereal crops and legumes, and the CWR of vegetables are still inadequately represented in these collections; only 5% of the accessions in global genebanks correspond to wild relatives of vegetable crops (FAO, 2010). Vegetables are an important source of nutrients, antioxidants and bioactive compounds and are essential for a healthy diet and for reducing the risk of some major chronic diseases (Slavin and Lloyd, 2012).

Eggplant, also known as brinjal or aubergine (*Solanum melongena* L., Solanaceae), is an important vegetable crop cultivated worldwide, and is the second most important solanaceous fruit crop after tomato (*S. lycopersicum* L.) (FAOSTAT, 2015). Because of its importance for food security eggplant is included with 34 other crops in the Annex 1 of the International Treaty on Plant Genetic Resources for Food and Agriculture (Fowler *et al.*, 2003; FAO, 2009), a mechanism established to facilitate access and use of plant genetic resources. Unlike tomato and potato (*S. tuberosum* L.), eggplants have their origins in the Old World (Weese and Bohs, 2010). Three distinct cultivated species of eggplants are known: the Gboma eggplant (*S. macrocarpon* L.), the scarlet eggplant (*S. aethiopicum* L.), and the brinjal eggplant itself (*S. melongena*). While the two former are mainly grown in small scale local agricultural systems in Africa (Daunay and Hazra, 2012), the eggplant is cultivated worldwide, with focus in Mediterranean and Asia in particular, where it ranks among the top five most important vegetable crops (Frary *et al.*, 2007). In this article we focus on the brinjal eggplant or aubergine, commonly referred to and referred to hereafter as the eggplant.

Eggplant is an important source of antioxidants, vitamins, and minerals (Raigón *et al.*, 2008; Gramazio *et al.*, 2014). The most important nutritional components of eggplant are phenolic compounds that give the fruit its antioxidant properties (Plazas *et al.*, 2013) which are beneficial for a number of metabolic and cardiovascular ailments. Breeding objectives for eggplant are mostly oriented toward developing high-yielding, early maturing, high fruit quality varieties, with stress resistance and high antioxidants (Singh and Kumar, 2007; Daunay and Hazra, 2012; Gramazio *et al.*, 2014). Breeders have performed research with wild relatives of eggplant for yield increase, fruit quality, disease resistance, and more recently, improved nutritional content (Table 4.1). However, in contrast to tomato, where most modern cultivars carry genes introgressed from CWR (Díez and Nuez, 2008), at present no modern commercial cultivars of eggplant with traits introgressed from CWR have been released.

All three cultivated eggplants are members of the diverse genus *Solanum* L., which is one of the largest genera of flowering plants with more than 1,000 species distributed worldwide (Bohs, 2005; Särkinen *et al.*, 2013). All eggplants belong to the Leptostemonum clade, a monophyletic group informally called the 'spiny' solanums, and within Leptostemonum, a species-rich clade composed exclusively of Old World taxa (the Old World clade sensu Stern *et al.*, 2011; Vorontsova *et al.*, 2013) from Africa, Australia, and Asia (including Eurasia and the Middle East).

The putative progenitor of *S. melongena* is *S. insanum* L., a species widespread in tropical Asia from Madagascar to the Philippines. The eggplant was domesticated in Asia; Vavilov (1951) included eggplant in his Indo-Burman center of diversity but more recently authors have

Table 4.1: Spiny solanums used in eggplant breeding and improvement programmes

Species	**Reported use**	**References**
S. anguivi Lam.	Backcross to obtain cytoplasm substitution male-sterile lines	Khan and Isshiki (2009)
S. incanum L.	Interspecific hybrids as rootstock	Gisbert *et al.* (2011)
	Backcross introgression of high content in bioactive phenolic acids	Prohens *et al.* (2013)
	Resistance to *Fusarium* wilts.	Rao and Kumar (1980)
S. linnaeanum Hepper & P.-M.L.Jaeger	Backcross introgression of resistance to *Verticillium dahliae*.	Sunseri *et al.* (2003); Liu *et al.* (2015)
	Development of introgression lines	Mennella *et al.* (2010)
S. sisymbriifolium Lam.	Sexual and somatic hybridization for resistance to *Ralstonia solanacearum* and *Verticillium dahliae*. No backcrosses or selfings of hybrids obtained.	Bletsos and Roupakias (1998); Collonnier *et al.* (2003a)
	Resistance to *Ralstonia solanacearum* and *Verticillium dahliae* wilts	Collonnier *et al.* (2003a)
	Resistance to fruit and shoot borers (*Leucinodes orbonalis*)	Chelliah and Srinivasan (1983)
	Resistance to root-knot nematodes (*Meloidogyne* spp.)	Ahuja *et al.* (1987)
	Resistance to spider mite	Schalk *et al.* (1975)
S. violaceum Ortega	Selection of selfings of interspecific hybrids resistant to *Fusarium* wilt	Rao and Kumar (1980)
	Backcross to obtain cytoplasm substitution male-sterile lines	Khan and Isshiki (2009)
S. torvum Sw.	Sexual and somatic hybridization for resistance to *Ralstonia solanacearum*, *Verticillium dahliae* and root-knot-nematodes. No backcrosses or selfings of hybrids obtained	Bletsos and Roupakias (1998); Jarl *et al.* (1999); Kumchai *et al.* (2013)
	Resistance to soil pathogens *R. solanacearum* and *Verticillium* and as graftstock	Jarl *et al.* (1999); Collonnier *et al.* (2003b); Gisbert *et al.* (2011)
	Resistance to root-knot nematodes (*Meloidogyne* spp.)	Ahuja *et al.* (1987)

suggested it could have two centers of domestication in China and/or India (e.g., Daunay and Janick, 2007; Wang *et al.*, 2008; Weese and Bohs, 2010), with an additional and independent center of domestication in the Philippines (Meyer *et al.*, 2012a). Eggplant is one of only a handful of crops that are postulated to have multiple centers of domestication (Meyer *et al.*, 2012b). *Solanum insanum* and *S. melongena* share pollinators and freely interbreed where cultivated and wild forms come together (Meyer *et al.*, 2012a; Davidar *et al.*, 2015). Wild populations are distinct from cultivated forms (Meyer *et al.*, 2012a) and in general harbor higher levels of genetic diversity (Mutegi *et al.*, 2015).

The rest of the close wild relative species of eggplant occur outside the main centers of eggplant domestication, mostly in tropical eastern Africa and the Middle East. Recent taxonomic studies have clarified the species identities and distributions of this complex set of taxa (Knapp *et al.*, 2013; Vorontsova and Knapp, in press), and phylogenetic analyses have clarified their relationships. Previous classification systems (Lester and Hasan, 1991; Daunay and Hazra, 2012) for these taxa used two species names only (*S. melongena* and *S. incanum*) with a set of groups within each largely defined by geography and delimited by letters (e.g., *S. incanum* group D for what Knapp *et al.* (2013) defined as *S. lichtensteinii* Willd.). This nomenclature complicated recording of species identities in genebanks with many accessions being identified only to species level (for a complete explanation of this system see Knapp *et al.*, 2013)

Previous work in analysing the relationships of eggplant wild relatives used only a small set of taxa that were morphologically similar to the cultivated species (e.g., Lester and Hasan, 1991). Molecular phylogenetic studies confirmed the close relationship of these species in what was termed the Eggplant clade, but showed that additional species (e.g., *S. linnaeanum* Hepper & P.-M.L.Jaeger) were also members of the monophyletic group (Weese and Bohs, 2007). Further studies using a large number of African and Asian species of spiny solanums (Vorontsova *et al.*, 2013; Aubriot *et al.*, 2016) confirmed the monophyly of the Eggplant clade, and also identified several poorly resolved groups that were also possible CWR of eggplants. One of these was called the Anguivi grade by Vorontsova *et al.* (2013) and comprised a large number of African and Asian species whose relationships were poorly resolved, including the other species of cultivated African eggplants *S. aethiopicum* and *S. macrocarpon* (see Figure 4 in Vorontsova *et al.* (2013)). The re-evaluation of species limits in eggplant wild relatives in the broadest sense coupled with studies further resolving the phylogenetic relationships of Old World spiny solanums in general means these CWR can be set in both a taxonomic and evolutionary framework that allows for progress to be made in conservation analyses.

Although CWR have some previous use in eggplant breeding, the comprehensiveness of the conservation and availability of these species in genebanks worldwide has not been assessed. Here we use gap analysis and threat assessment methods to identify gaps in genebank collections, designate those species and geographic areas most threatened and least well conserved *ex situ*, and propose actions to overcome these deficiencies.

4.3 Materials and methods

4.3.1 Gene pool concept and selection of species

The gene pool concept is used in plant breeding to delineate species based on their ability to cross with the domesticated species, and on the relative fertility of the offspring (Harlan and de Wet, 1971). Because crossability relationships have not been established for numerous crops and their associated wild species, surrogate categories have been created based on taxonomic hierarchy (Maxted *et al.*, 2006), and a combination of evolutionary relationships, ploidy level and reproductive biology (Wiersema *et al.*, 2012). For eggplant, or other crops that are members of large genera such as *Solanum*, the delimitation of biologically meaningful gene pools is challenging due to limited crossability data that are systematically reported (Plazas *et al.*, 2016) and the extremely large number of potential gene pool members. Here we adopt a combined approach in order to delimit a meaningful gene pool concept for the cultivated eggplant taking into account both relatedness as measured by phylogenetic analyses (Vorontsova *et al.*, 2013; Aubriot *et al.*, 2016) and, where available, crossability data (Plazas *et al.*, 2016).

4.3.2 Occurrence data

We collected occurrence data for the species we defined as wild relatives of the brinjal eggplant (Table 4.2). Solanaceae Source (`http://solanaceaesource.org`) was used as the principal source of occurrence records derived from herbarium specimen data, because it holds information for taxonomically verified specimens used in recent monographic work and recent field collections (Vorontsova and Knapp, in press). All specimens containing sufficient locality data were manually georeferenced to a high standard and those previously georeferenced by others were checked. Additional herbarium specimen data were retrieved from the Global Biodiversity Information Facility (GBIF; `http://www.gbif.org`) and were similarly checked for nomenclatural and geographic accuracy. Herbarium accessions are mapped in Figure 4.1. Genebank accession passport data from multiple genebanks was electronically retrieved from the Global Information Portal for Plant Genetic Resources (Genesys; `https://www.genesys-pgr.org`), and through direct requests to genebanks holding collections of eggplants and eggplant wild relatives. All data providers of both herbarium and germplasm accessions are listed in Supplementary Table 4.1.

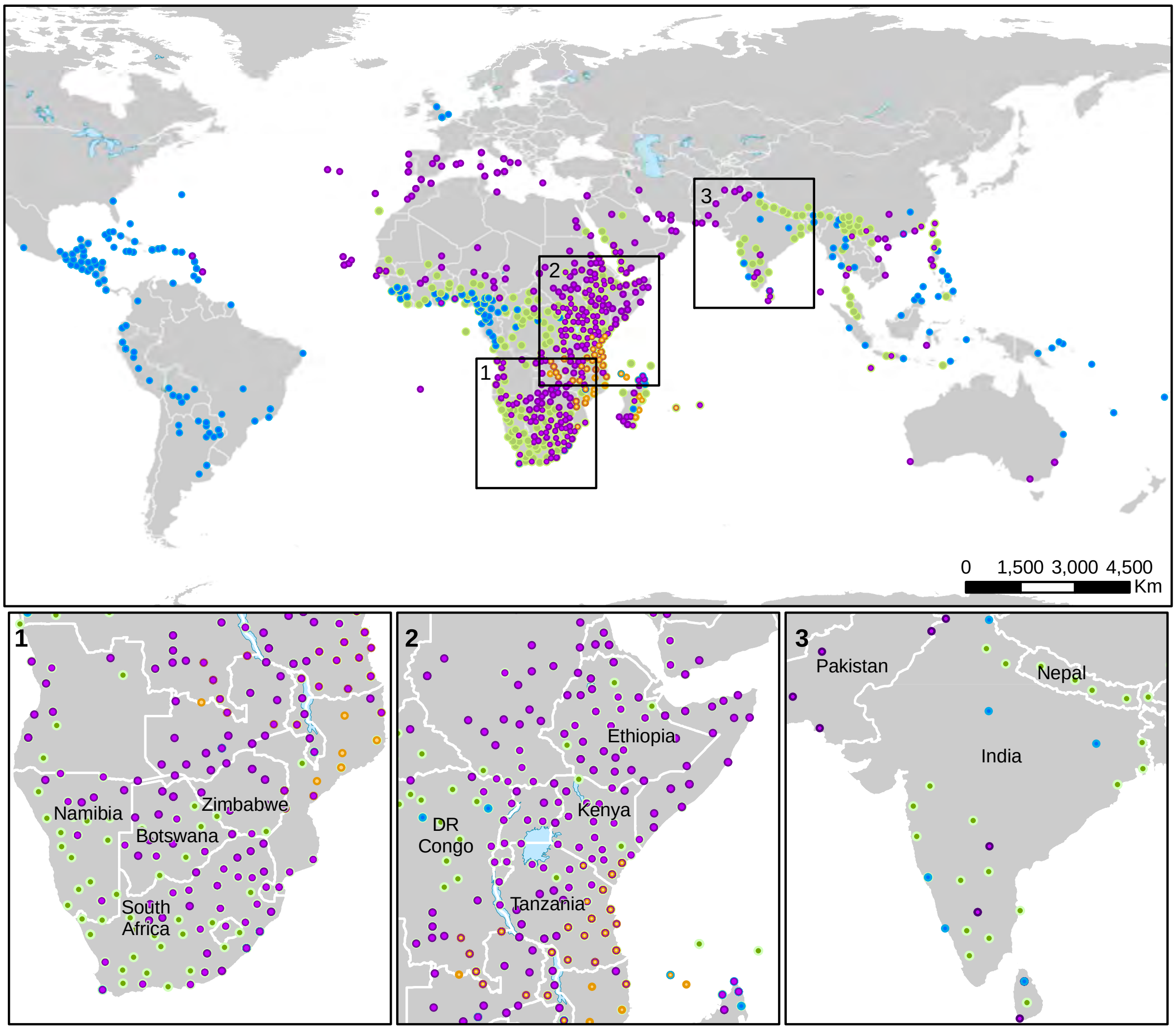


Figure 4.1: Map of herbarium specimens with geographical coordinates of spiny solanums (Leptostemonum Clade) used in this study by phylogenetic group: purple points= Eggplant clade; green points= Anguivi grade; orange points= Climbing clade; blue points= New World species.

4.3.3 Species distribution modelling

Georeferenced herbarium records were used as occurrence samples to build species distribution models (SDMs) for species with more than five (5) unique georeferenced records (Wisz *et al.*, 2008). Due to taxonomic uncertainties in genebank records, only verified herbarium specimen data were used to construct SDMs. All available environmental data layers were obtained from the Worldclim database, version 1.4 (`http://www.worldclim.org`; Hijmans *et al.*, 2005), along with potential evapotranspiration (PET), actual evapotranspiration (AET), and aridity from the CGIAR Consortium for Spatial Information (CGIAR-CSI; `http://www.cgiar-csi.org`) and enhanced vegetation index (EVI) derived from MODIS imagery from WorldGrids (`http://www.worldgrids.org`). All the data layers were obtained at a spatial resolution of 30 arc second ($\sim$ 1 km at the Equator). Variable selection was based on a combination of cluster, correlation, and principal components analyses as well as expert judgement based on the ecology of the species. Our final variable selection was driven by choosing variables that were ecologically relevant for eggplant wild relatives, following recommendations for best practice (Elith and Leathwick, 2009). The final selection of six environmental variables were: temperature seasonality, annual mean temperature, the precipitation of the coldest quarter (three month period), enhanced vegetation index (EVI), actual evapotranspiration (AET) and water deficit (calculated as potential evapotranspiration minus actual evapotranspiration; Stephenson, 1998).

MaxEnt (Version 3.3.3; Phillips *et al.*, 2006) was used to build species distribution models as it has been found to be among the best-performing of the various correlative SDM approaches available, and because it is able to perform well with small samples of presence localities (Elith *et al.*, 2006; Pearson *et al.*, 2007). We adopted the default regularization parameters but restricted MaxEnt to using only linear and quadratic functional forms, which constrains models to produce relatively simple models that do not over-fit to the training data (Merow *et al.*, 2013; Syfert *et al.*, 2013). Sampling bias was controlled by including a sampling bias dataset (Phillips *et al.*, 2009) constructed from all georeferenced plant occurrence data from both the GBIF data portal and all georeferenced records from Solanaceae Source (6,200,000 and 47,500 records respectively; downloaded 24 September 2014). The spatial extent of the area modelled also influences model performance (Peterson *et al.*, 2011). In our case, species ranges varied from very small within a single country (e.g., *S. deflexicarpum* C.Y.Wu & S.C.Huang) to widespread on multiple continents (e.g., *S. torvum* Sw.). Given the influence of the spatial extent of the area modelled on model performance, we built models with varied extents that fit species ranges in order to allow for a biologically meaningful fit between a species occurrence and the associated environmental variables. A 200 km buffer around all presence points was used to delimit the model extent of each species following VanDerWal *et al.* (2009). All species with >10 georeferenced samples were modelled using a 5-fold cross-validation approach to assess model predictive accuracy. This approach uses 80% of the data to train the model and reserves 20% for model evaluation; this process was repeated until each reserved set had been used to evaluate models (Franklin, 2009). A leave-one-out method was performed for species with sample sizes from 5 to 10 data points (Pearson *et al.*, 2007). SDM performance was evaluated by using the area under the curve (AUC) in a receiver operating characteristic (ROC) plot; an AUC value of 1.0 indicates perfect discrimination ability and a value of 0.5 or less indicates a prediction no better than random. Only models with an average test AUC above 0.7 were used for further analyses. In the case that the SDM did not have an average test AUC above 0.7, a convex hull was created and used in place of the SDM for the rest of the *ex situ* gap analysis. For species with AUC above 0.7, SDMs were then reconstructed using the complete set of occurrences to develop a more robust predictive model (Fielding and Bell, 1997) and these models were used for further analyses. Binary maps of predicted presences and absences were created for each species using a maximum sensitivity and specificity threshold (Liu *et al.*, 2005).

4.3.4 *Ex situ* conservation analysis

We used three species-specific gap analysis metrics adapted from Ramírez-Villegas *et al.* (2010) to estimate the sufficiency of accessions currently conserved in genebanks in representing the full diversity of CWR species based on their potential distributions and the known wild diversity by using data derived from herbarium collections; the sampling representativeness score (SRS), geographic representativeness score (GRS), and ecological representativeness score (ERS). Each metric has a scale from 0 to 10. These metrics measure the representativeness of the germplasm collections in relation to the modelled distribution of each crop wild relative species, and the known wild diversity using data derived from herbarium collections. The SRS is a general measure of how well germplasm records represent the known wild diversity in terms of numbers of records, and is calculated by dividing the number of germplasm records by the total number of records (germplasm + herbarium). An advantage of the SRS metric is that it permits the use of all records, regardless of whether they have geographical coordinates. The GRS focuses on the geographic representativeness of germplasm collections, and is measured by comparing the current geographic coverage of the germplasm records calculated with a 50 km circular buffer surrounding the original site of collections of each germplasm accession record (Hijmans and Spooner, 2001) against the SDM of each species. The ERS measures ecological representation of germplasm collections, and is calculated by comparing the number of terrestrial ecoregions (Olson *et al.*, 2001) represented in the geographic coverage of the germplasm collections (buffer of 50 km radius around each collection) against the total number of ecoregions encompassed in the SDM of each species. The metrics were averaged to produce a final priority score (FPS), which was then classified into four categories of urgency for further collecting in order to improve representation in genebank collections: (1) high priority for further collecting (FPS = 0-3); (2) medium priority (FPS = 3.01-5); (3) low priority (FPS = 5.01-7.5; and (4) sufficiently represented (FPS = 7.51-10). Species with no germplasm accessions conserved in genebanks (with or without geographic coordinates) received a final score of 0 and were listed as high priority species.

We produced maps to identify areas prioritized for further collecting of eggplant CWR seeds by excluding the original sites of collection of existing germplasm accessions with coordinates (50 km buffer around each record) from the geographic coverage derived from SDMs. When an appropriate SDM was not produced, 50 km circular buffers around each herbarium record were used as proxies for geographic coverage. Remaining geographic areas represent areas where a species is potentially present but has not yet been collected (i.e., collecting gaps). Individual maps of species from the high and medium priority categories were added together to produce a priority collecting hotspots map. Maps illustrating collecting hotspots were created for the following groups of species: 1) Eggplant clade, 2) Anguivi grade, 3) Climbing clade, and 4) the New World species.

All gap analyses were performed in R v2.15 using maptools (Bivand and Lewin-Koh, 2014), rgdal (Bivand *et al.*, 2014), SDMTools (VanDerWal *et al.*, 2014), raster (Hijmans, 2014), sp (Pebesma and Bivand, 2005; Bivand *et al.*, 2013) and dismo (Hijmans *et al.*, 2014) packages. All raster files were mapped in ESRI ArcGIS 10 (ESRI, 2011).

4.3.5 *In situ* conservation assessment

We evaluated the preliminary conservation status of the wild relatives of eggplant using the IUCN Red List Categories and Criteria (IUCN, 2012), one of the most widely used systems for assessing the conservation status of species in their natural habitats (Rodrigues *et al.*, 2006; Mace *et al.*, 2008). Threatened species were placed within three categories (critically endangered, endangered and vulnerable) based on five criteria (A-E; IUCN, 2012). A fourth category, near threatened, was used when species partially met the IUCN threat criteria and were present in areas with documented rapid change in land use. Because population data were lacking for the eggplant wild relatives, several of the Red List criteria were inapplicable and we focused on using estimates of geographic range as the key determinant of extinction risk (criterion B). The extent of occurrence (EOO), which measures the overall geographic spread of the localities in which a species is found (Gaston and Fuller, 2009), was calculated by applying a minimum convex polygon (MCP; also known as a convex hull) around the known species locations (IUCN, 2012). Area of occupancy (AOO), the other geographical range element to criterion B, measures the area with in the EOO that is occupied by a taxon. We focused on EOO rather than AOO because not enough information was available from herbarium specimen records to know whether populations were well sampled within the species range. AOO calculations were only performed for those species with a preliminary status indicating some degree of threat. The effect of land use changes for each species was assessed in Google Earth using current and historical land use images to visually determine fragmentation within the EOO due to human activities such as mining, agriculture and urban habitation. Following IUCN (2012) species were considered threatened if the EOO was < 20,000 km^2 (< 100 km^2 = critically endangered, < 2,000 km^2 = endangered, < 20,000 km^2 = vulnerable) and if a minimum of two of the following subcategories were met: severely fragmented habitat, continuing decline of habitat, and extreme range fluctuations (see Supplementary Table 4.2 for details of IUCN criteria).

We also examined the intersection of threatened species of eggplant CWR and protected areas in order to further refine our prioritization of collecting and conservation action. First we focused on regions with priority collecting hotspots (i.e., those in which more than seven under-collected species were estimated to occur in the same pixel) and evaluated regions within and outside of protected areas as defined from the World Database on Protected Areas (World Database on Protected Areas Consortium, 2010). We also prepared a map of hotspots for further collecting of those species preliminarily assessed at risk of extinction by overlaying the collecting priority map for each species identified in one of the threatened categories (including those in the near threatened category). To further prioritize regions for conservation actions, the extent of area in collecting hotspots for species at risk of extinction was combined with the overall area for general priority collecting hotspots to identify regions with multiple species poorly represented in genebanks that coincide with regions where there are also significant conservation threats.

4.4 Results

4.4.1 Gene pool concept definition

Based on the existing phylogenetic hypotheses and information on crosses between the wild relatives and cultivated eggplant, we included 52 eggplant CWR species in our eggplant wild relative gene pool concept (Table 4.2). Species were assigned to clades, grades, and gene pools based on morphological evidence when they had not been included in a previous phylogenetic analysis (Table 4.2; Vorontsova and Knapp, in press). Gene pool 1 (GP1) was defined as the cultivated eggplant and its widely accepted wild progenitor *S. insanum* (Knapp *et al.*, 2013), following Maxted *et al.* (2006) concept. Gene pool 2 (GP2) was defined based on phylogenetic evidence from Vorontsova *et al.* (2013), where the eggplant and its closest relatives are resolved in a large monophyletic clade, that includes three elements: (a) the monophyletic Eggplant clade that includes the most closely related species (excluding *S. insanum* and *S. melongena* itself), (2) the small Climbing clade, and (3) the set of species from the paraphyletic Anguivi grade where relationships amongst the component species are not well resolved. Lastly, gene pool 3 (GP3) is defined to include phylogenetically more distantly related species native to the New World which have already been used in eggplant breeding due their resistance traits (e.g., *S. sisymbriifolium* Lam., *S. torvum*). The Eggplant clade included 10 species, the Climbing clade three species, the Anguivi grade 36 species, and the three remaining species were included in a non-phylogenetically defined group we here call the New World group (Table 4.2). *Solanum anguivi* Lam. (wild ancestor of cultivated *S. aethiopicum*) and *S. dasyphyllum* Schumach. & Thonn. (wild ancestor of cultivated *S. macrocarpon*) are included in our analysis as members of the Anguivi grade.

Table 4.2: Eggplant wild relative species with gap analysis results priority categories, and preliminary IUCN Red List status. Species are ordered alphabetically by epithet. Where species have not been included in molecular phylogenetic analyses, relationships are inferred based on morphological similarity (Vorontsova and Knapp, in press). Species where species distribution models (SDMs) were not constructed due to lack of georeferenced records are indicated with an asterisk (*), and species where models failed due to low AUC values (<0.7) are labelled with a cross (+). 1 = Species native to the small Aldabra islands in the Indian Ocean, for which no global environmental data recorded and hence no SDM calculated, 2 = Values excluding Caribbean range where species is likely extinct, 3 = Species only known from the type. SRS = Sampling Representativeness Score, GRS = Geographic Representativeness Score, ERS = Ecological Representativeness Score, FPS = Final Priority Score, ECAT = *Ex situ* conservation prioritization category. Samples available for SDMs were based on georeferenced herbarium records and duplicates within a 1km^2 were removed.

Species	**Phylogenetic group (gene pool)**	**Native distribution**	**No. of herbarium records (with coordinates)**	**No. of germplasm records (with coordinates)**	**No. of samples available for SDMs**	**Training AUC (Test AUC)**	**SRS**	**GRS**	**ERS**	**FPS**	**ECAT**
S. adoense Hochst. ex A.Rich.	Anguivi (GP2)	Africa	78 (36)	0 (0)	36	0.769 (0.708)	0	0	0	0	HPS
S. agnewiorum Voronts.	Eggplant (GP2)	Africa	12 (8)	0 (0)	9	0.89 (0.87)	0	0	0	0	HPS
S. aldabrense C.H.Wright*,[1]	Anguivi (GP2)	Aldabra (Seychelles)	30 (10)	2 (2)	10	- (-)	0.6	0	0	0.6	HPS
S. anguivi Lam.	Anguivi (GP2)	Africa and Madagascar	717 (514)	115 (46)	514	0.747 (0.743)	1.4	0.6	8.7	3.6	MPS
S. aureitomentosum Bitter+	Eggplant (GP2)	Africa	29 (19)	0 (0)	19	- (-)	0	0	0	0	HPS
S. burchellii Dunal	Anguivi (GP2)	Africa	36 (31)	5 (4)	32	0.896 (0.867)	1.2	0.4	4.2	1.9	HPS
S. campylacanthum Hochst. ex A.Rich.+	Eggplant (GP2)	Africa	1101 (715)	66 (48)	715	- (-)	0.6	0.2	7.1	2.6	MPS
S. capense L.	Anguivi (GP2)	Africa	58 (45)	22 (12)	45	0.795 (0.784)	2.8	0.7	5.7	3.1	MPS
S. catombelense Peyr.	Anguivi (GP2)	Africa	64 (60)	6 (5)	60	0.817 (0.77)	0.9	0.2	3.7	1.6	HPS
S. cerasiferum Dunal	Eggplant (GP2)	Africa	64 (45)	5 (3)	45	0.832 (0.801)	0.7	0	1.3	0.7	HPS
S. cyaneopurpureum De Wild.	Anguivi (GP2)	Africa	36 (22)	1 (1)	22	0.777 (0.723)	0.3	0.2	2.1	0.9	HPS
S. dasyphyllum Schumach. & Thonn. +	Anguivi (GP2)	Africa	209 (174)	35 (19)	174	- (-)	1.4	0	4	1.8	HPS
S. deflexicarpum C.Y.Wu & S.C.Huang*	Anguivi (GP2)	China	3 (2)	0 (0)	2	- (-)	0	0	0	0	HPS
S. glabratum Dunal	Anguivi (GP2)	Arabia	132 (79)	0 (0)	79	0.799 (0.779)	0	0	0	0	HPS
S. hastifolium Hochst. ex Dunal	Anguivi (GP2)	Africa	130 (95)	1 (0)	95	0.843 (0.829)	0.1	0	0	0	HPS
S. hovei Dunal	Anguivi (GP2)	India	12 (9)	0 (0)	8	- (-)	0	0	0	0	HPS
S. humile Lam.	Anguivi (GP2)	South Africa	44 (35)	15 (12)	35	0.889 (0.874)	2.5	0.4	5	2.7	MPS

Table 4.2 – continued from previous page

Species	Phylogenetic group (gene pool)	Native distribution	No. of herbarium records (with coordinates)	No. of germplasm records (with coordinates)	No. of samples available for SDMs	Training AUC (Test AUC)	SRS	GRS	ERS	FPS	ECAT
S. inaequiradians Werderm.	Anguivi (GP2)	Africa	9 (7)	0 (0)	7	0.915 (0.838)	0	0	0	0	HPS
S. incanum L.	Eggplant (GP2)	Africa and Middle East	329 (184)	114 (37)	184	0.881 (0.878)	2.6	0.1	5.3	2.7	MPS
S. insanum L.	Eggplant (GP1)	Southeast Asia and Madgascar	173 (78)	9 (9)	78	0.871 (0.847)	0.5	0.1	1.7	0.7	HPS
S. lamprocarpum Bitter	Anguivi (GP2)	Africa	9 (9)	1 (0)	9	- (-)	1	0	0	0.5	HPS
S. lichtensteinii Willd.	Eggplant (GP2)	South Africa	90 (71)	15 (14)	71	0.789 (0.772)	1.4	0.3	3.8	1.8	HPS
S. lidii Sunding*	Anguivi (GP2)	Canary Islands	7 (3)	14 (4)	3	- (-)	6.7	10	10	8.9	NFCR
S. linnaeanum Hepper & P.-M.L.Jaeger	Eggplant (GP2)	South Africa; Mediterreanean	142 (110)	75 (28)	110	0.939 (0.93)	3.5	0.6	2.9	2.3	HPS
S. litoraneum A.E.Gonç.	Anguivi (GP2)	Africa	4 (4)	0 (0)	4	0.88 (0.85)	0	0	0	0	HPS
S. macracanthum A.Rich.	Anguivi (GP2)	Africa	29 (19)	0 (0)	19	0.881 (0.785)	0	0	0	0	HPS
S. malindiense Voronts.	Anguivi (GP2)	Africa	16 (10)	0 (0)	10	0.966 (0.964)	0	0	0	0	HPS
S. mauense Bitter	Anguivi (GP2)	Africa	47 (42)	1 (0)	42	0.87 (0.854)	0.2	0	0	0.1	HPS
S. multiflorum Roth*	Anguivi (GP2)	India	13 (5)	0 (0)	5	- (-)	0	0	0	0	HPS
S. nigriviolaceum Bitter	Anguivi (GP2)	Africa	33 (25)	3 (2)	32	0.892 (0.877)	0.8	10	10	6.9	LPS
S. platacanthum Dunal	Anguivi (GP2)	Arabia	18 (14)	1 (1)	14	- (-)	0.5	7.6	10	6	LPS
S. polhillii Voronts.	Anguivi (GP2)	Africa	56 (30)	0 (0)	25	0.912 (0.881)	0	0	0	0	HPS
S. richardii Dunal	Climbing (GP2)	Africa and Madagascar	119 (77)	3 (2)	77	0.725 (0.677)	0.2	0	10	3.4	MPS
S. rigidum Lam.[2]	Eggplant (GP2)	Cape Verde Islands	33 (32)	0 (0)	25	0.967 (0.963)	0	0	0	0	HPS
S. rubetorum Dunal	Anguivi (GP2)	Africa	56 (38)	9 (1)	30	0.851 (0.81)	1.4	0.2	2.1	1.2	HPS
S. ruvu Voronts.*[3]	Anguivi (GP2)	Africa	1 (1)	0 (0)	1	- (-)	0	0	0	0	HPS
S. setaceum Dammer	Anguivi (GP2)	Africa	39 (29)	0 (0)	29	0.79 (0.75)	0	0	0	0	HPS
S. sisymbriifolium Lam.	NW-Sisymbrifolium (GP3)	South America (native); widespread	151 (109)	68 (23)	109	0.806 (0.778)	3.1	0	2.8	2	HPS
S. sodomeodes Kuntze	Anguivi (GP2)	Africa	21 (21)	0 (0)	21	0.841 (0.783)	0	0	0	0	HPS
S. stipitatostellatum Bitter	Climbing (GP2)	Africa	56 (34)	0 (0)	38	0.872 (0.8)	0	0	0	0	HPS
S. supinum Dunal	Anguivi (GP2)	Africa	24 (19)	5 (2)	19	0.854 (0.816)	1.7	0.1	0.6	0.8	HPS
S. taitense Vatke	Anguivi (GP2)	Africa	43 (33)	0 (0)	33	0.819 (0.744)	0	0	0	0	HPS
S. tomentosum L.+	Anguivi (GP2)	South Africa	56 (25)	32 (17)	34	- (-)	3.6	1.4	9.2	4.7	MPS
S. torreanum Gonç.	Anguivi (GP2)	Africa	11 (10)	0 (0)	10	- (-)	0	0	NA	0	HPS

Table 4.2 – continued from previous page

Species	**Phylogenetic group (gene pool)**	**Native distribution**	**No. of herbarium records (with coordinates)**	**No. of germplasm records (with coordinates)**	**No. of samples available for SDMs**	**Training AUC (Test AUC)**	**SRS**	**GRS**	**ERS**	**FPS**	**ECAT**
S. torvum Sw.	NW-Torva (GP3)	Central America (native?); widespread	821 (540)	125 (55)	540	0.815 (0.816)	1.3	0.1	4.5	2	HPS
S. umtuma S.Knapp & Voronts.	Eggplant (GP2)	South Africa	22 (16)	0 (0)	16	0.789 (0.765)	0	0	0	0	HPS
S. usambarense Bitter & Dammer	Anguivi (GP2)	Africa	36 (32)	0 (0)	31	0.825 (0.785)	0	0	0	0	HPS
S. usaramense Dammer	Anguivi (GP2)	Africa	23 (15)	0 (0)	15	0.891 (0.89)	0	0	0	0	HPS
S. vespertilio Aiton	Anguivi (GP2)	Canary Islands	19 (9)	18 (8)	9	0.812 (0.751)	4.9	10	5	6.6	LPS
S. viarum Dunal	NW-Acanthophora (GP3)	Brazil (native); widespread	46 (34)	85 (45)	34	0.901 (0.879)	6.5	0.1	4.2	3.6	MPS
S. violaceum Ortega	Anguivi (GP2)	India and Southeast Asia	237 (108)	84 (41)	108	0.792 (0.776)	2.6	0.2	4.6	2.5	HPS
S. zanzibarense Vatke	Climbing (GP2)	Africa	74 (58)	1 (1)	58	0.87 (0.845)	0.1	0.1	1.5	0.6	HPS

4.4.2 Occurrence data

A total of 5,618 occurrence records from herbarium specimens and 936 germplasm accession records were compiled. Records with geographic coordinates ranged from 1 to 1,100 per species (Table 4.2). All herbarium specimen records for the eggplant gene pools defined above are mapped in Figure 4.1.

4.4.3 Species distribution models

SDMs with adequate predictive performance to use in the gap analysis were produced for 37 CWR species [median test AUC was 0.83 ± 0.06 SD (from 5-fold cross-validated models)]. SDMs could not be built for five species with fewer than five georeferenced records (*S. deflexicarpum*, *S. lidii* Sunding, *S. litoraneum* A.E.Gonç., *S. ruvu* Voronts.). A further seven species were not included even though they had slightly higher record numbers (5-25) because AUC values were below our targeted threshold of 0.7 (Table 4.2). While it is expected that sample sizes of less than 25 would yield SDMs with poor predictive performance, we also found that three species with very large numbers of records (*S. campylacanthum* Hochst. ex A.Rich. [n=715], *S. dasyphyllum* [n=174] and *S. richardii* Dunal [n=77]) also yielded SDMs with poor predictive performance. *Solanum campylacanthum* and *S. dasyphyllum* are both very widely distributed in many different habitats in continental Africa, and it is likely that climatic factors are not of primary importance in determining their distributions. It is possible that species with poor SDM performance have distributions that are constrained to factors (e.g., dispersion or biotic interactions) beyond the variables included to build the SDMs. For instance, a recent field study in the Kenyan savannah found that *S. campylacanthum* abundance greatly decreased in plots where several mammals browsed on different parts of the plants (Pringle *et al.*, 2014), suggesting biotic interactions could be important factors influencing the distribution of this species.

4.4.4 *Ex situ* conservation analysis

The wild relatives of eggplant are currently exceedingly poorly represented in global germplasm collections (Table 4.2). The final priority scores (FPS) for the gap analysis classified 40 species as high priority for further collecting to improve the state of their conservation in genebanks, eight (8) as medium priority for further collecting to improve the state of their conservation in genebanks, three as low priority for further collecting and only one species as currently sufficiently represented in genebanks (Table 4.2). A total of 34 of the 52 species (65%) have five or fewer accessions publicly available in genebanks.

The sampling representativeness score (SRS) influenced the assessment of the low level of representation of eggplant CWR species in genebanks, with many species represented by very few genebank accessions (Supplementary Figure 4.1a). Island species with narrow distributions (*S. lidii* and *S. vespertilio*) were an exception, with high SRS (Table 4.2). Species like *S. campylacanthum* had the largest number of herbarium specimen records, but also one of the largest gaps with respect to the number of germplasm collections (1,101 herbarium records as opposed to 66 germplasm accessions). This mismatch results from the use of revised species names and delimitations for verified herbarium material (Vorontsova and Knapp, in press), and the outdated naming of accessions in genebanks with a previous set of species names (based on Lester and Hasan, 1991).

The low level of representation of the wild relatives of eggplant in genebanks was also strongly influenced by the inadequacy of geographic coverage of samples (i.e., geographic locations where germplasm has already been collected as compared to the total modelled distributions of eggplant CWR) (Supplementary Figure 4.1b). Forty seven of the 52 species used in our analyses had a geographic representativeness score (GRS) of less than 2 (Table 4.2). Only species with relatively restricted distributions had a GRS above 7.5 (*S. lidii*, *S. vespertilio* Aiton, *S. platacanthum* Dunal, and *S. nigriviolaceum* Bitter); these species were also assessed as relatively well-represented with respect to ecological regions (ERS; ERS ≥ 5).

Ecological representation (ERS scores) of eggplant wild relatives is somewhat higher than either the SRS or GRS, but still falls below the ideal (see Supplementary Figure 4.1c). Twenty four of the 52 specie analysed had an ERS of < 2. Some widespread species such as *S. anguivi*, *S. richardii*, and *S. tomentosum* L. displayed a relatively comprehensive ecological representativeness (ERS >8). Caution should be taken with this result however, the terrestrial ecoregions defined by Olson *et al.* (2001) are relatively crude, and more tightly defined vegetation types that more accurately represent the diversity of terrestrial biomes (e.g., Särkinen *et al.*, 2011) could significantly change this representation.

Regions identified as hotspots for further collecting of multiple medium and high priority species included eastern and southern Africa and Southeast Asia (Figure 4.2). Several areas along the Tanzania-Kenya border have as many as 10 species occurring in the same area and were identified as high or medium priority collecting areas (Figure 4.2, inset 2). Further hotspots include western South Africa (Figure 4.2, inset 1) and the region of the Western Ghats in India (Figure 4.2, inset 3).

These "secondary" hotspots are even clearer when the data are analysed by phylogenetic grouping. The Eggplant clade, which is comprised of the 10 species most closely related to the cultivated eggplant, has priority areas with the greatest number of overlapping species for further germplasm collecting in Kenya and Tanzania, and with a small area to the south in Zimbabwe (Supplementary Figure 4.2a). Western South Africa and Namibia is identified as an additional priority for further collecting for the Anguivi grade (Supplementary Figure 4.2b), and areas in coastal eastern Africa from Kenya south to Mozambique are priorities for the species of the Climbing clade (Supplementary Figure 4.2c).

4.4.5 *In situ* conservation assessment

Solanum lidii is the only species of the 52 eggplant wild relatives that has been formally assessed and included on the IUCN Red List (critically endangered; Rodríguez Delgado *et al.*, 2011). We have assessed an additional nine species as threatened (EOO < 20,000 km^2) according to preliminary Red List criteria due to small geographic ranges (Figure 4.3) and occurrence in areas with high habitat fragmentation and degradation (IUCN criterion B1ab (iii); Table 4.3). Our analysis shows that one additional species is likely to be critically endangered (*S.*

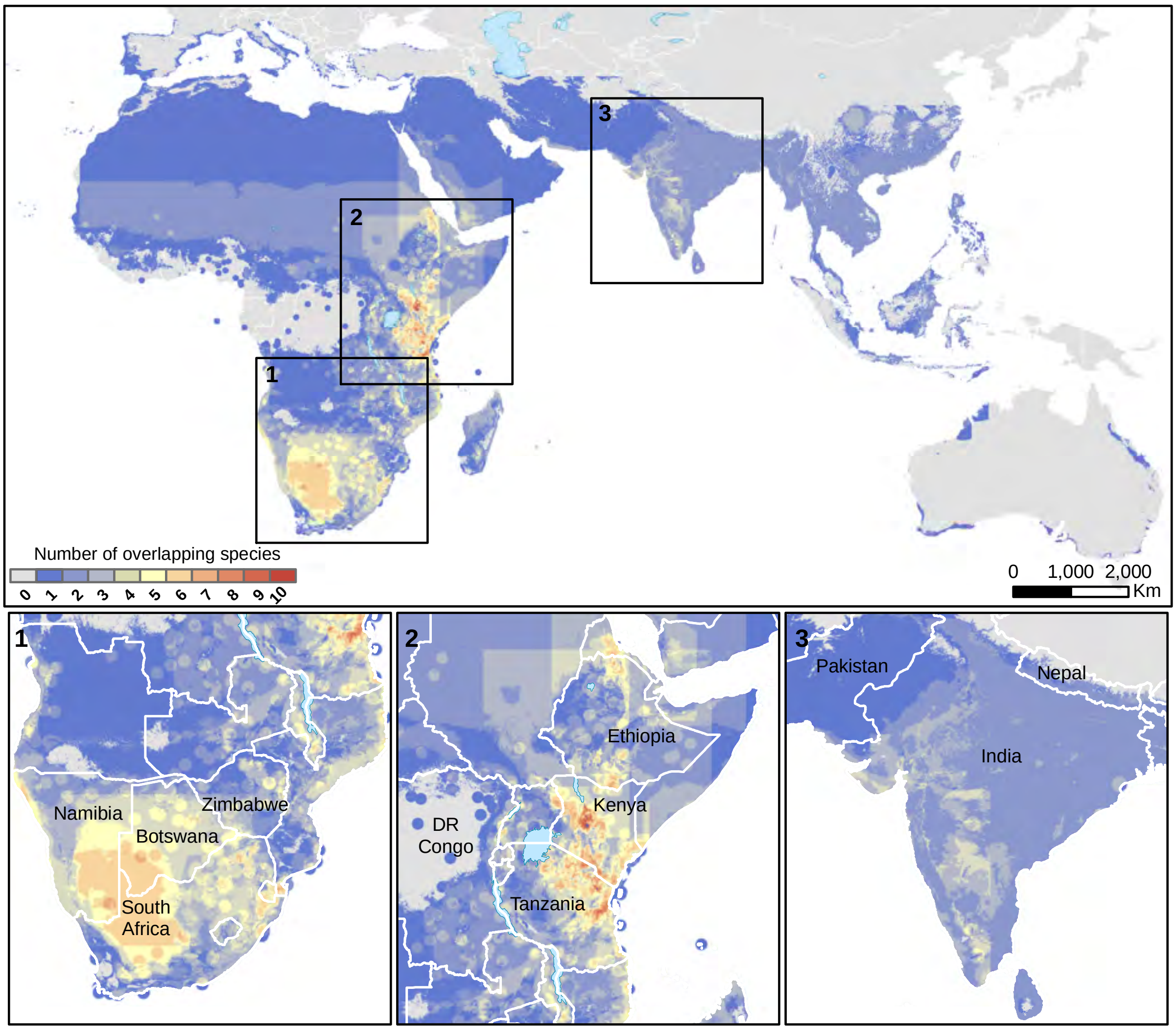


Figure 4.2: Map of future collecting hotspots for 48 species of cultivated eggplant wild relatives classified as medium or high priority based on the gap analysis. Areas with orange-red colour show highest priority collecting hotspots; these include more than seven species. Insets zoom in on regions identified as 1 (eastern Africa), 2 (southwestern Africa) and 3 (India) on the larger scale map.

deflexicarpum), four species endangered (*S. aldabrense* C.H.Wright, *S. inaequiradians* Bitter, *S. malindiense* Voronts., and *S. vespertilio*), and four species vulnerable (*S. agnewiorum* Voronts., *S. litoraneum*, *S. setaceum* Bitter, and *S. torreanum* A.E.Gonç.; Table 4.3). In addition to the 10 eggplant wild relatives assessed as threatened, two species with EOOs greater than 20,000 km^2 are likely to be near threatened (*S. platacanthum* and *S. nigriviolaceum*) based on their occurrence in regions with high land degradation. A third species, *S. rigidum* Lam., could also be considered to be near threatened because it has been recently shown to be endemic to the Cape Verde Islands although there are a few old specimens from the Caribbean, possibly dispersed by humans (Knapp and Vorontsova, 2013). The most poorly known species of the eggplant wild relatives is *S. ruvu* that is only known from the type and has been suggested to be extinct in the wild; recent collection trips to the type locality and similar habitats have failed to encounter extant populations (Vorontsova and Mbago, 2010). In total, one quarter of the species of eggplant CWR (14 spp.) are assessed to have some level of threat to their wild populations (see Table 4.3), while the rest (38 spp.) are assessed here as of least concern, meaning their populations are widespread and not of immediate conservation concern (see Table 4.3 for EOO and AOO values for all threatened species and Supplementary Table 4.3 for EOO values for all species).

Table 4.3: IUCN threat assessments for eggplant wild relatives at risk of extinction. Extent of Occurrence (EOO) and Area of Occupancy (AOO) calculations described in the text. All listings are new here with the exception of *S. lidii* (formally assessed by Rodríguez Delgado *et al.*, 2011). EOO and AOO values for all species are in Supplementary Table 4.3. [a]EOO could not be calculated because there are not only two distinct localities. [b]AOO value is as reported from Rodríguez Delgado *et al.* (2011). [c]Values excluding Caribbean range where species is likely to be not represent no longer.

Species	**EOO (km^2)**	**AOO (km^2)**	**Preliminary IUCN category**
S. agnewiorum Voronts.	7946	32	VU B1ab(iii)
S. aldabrense C.H.Wright	3437	32	EN B1ab(iii)
S. deflexicarpum C.Y.Wu & S.C.Huang[a]	0	8	CR B1ab(iii)
S. inaequiradians Werderm.	2101	28	EN B1ab(iii)
S. lidii Sunding[b]	<5	5	CR B2ab(ii,iii) + C2a(i)
S. litoraneum A.E.Gonç.	6196	16	VU B1ab(iii)
S. malindiense Voronts.	2381	40	EN B1ab(iii)
S. nigriviolaceum Bitter	43749	16	NT
S. platacanthum Dunal	45123	56	NT
S. rigidum Lam.	135,509 (35,310)[c]	124 (108)[c]	LC (NT)[c]
S. ruvu Voronts.	0	4	EX [?]
S. setaceum Dammer	13564	112	VU B1ab(iii)
S. torreanum Gonç.	16804	40	VU B1ab(iii)
S. vespertilio Aiton	2336	36	EN B1ab(iii)

Five of the species identified at risk of extinction are restricted to eastern Africa, primarily in Kenya, Tanzania, and Uganda (Figure 4.3). This is also the region identified as containing multiple species under-represented in *ex situ* collections, particularly along the border between Kenya and Tanzania (Figure 4.4a). Only 15% (ca. 10,000 km^2 of a total 65,000 k^2) of hotspot areas identified as high priority for further collecting (i.e., seven or more eggplant wild relatives

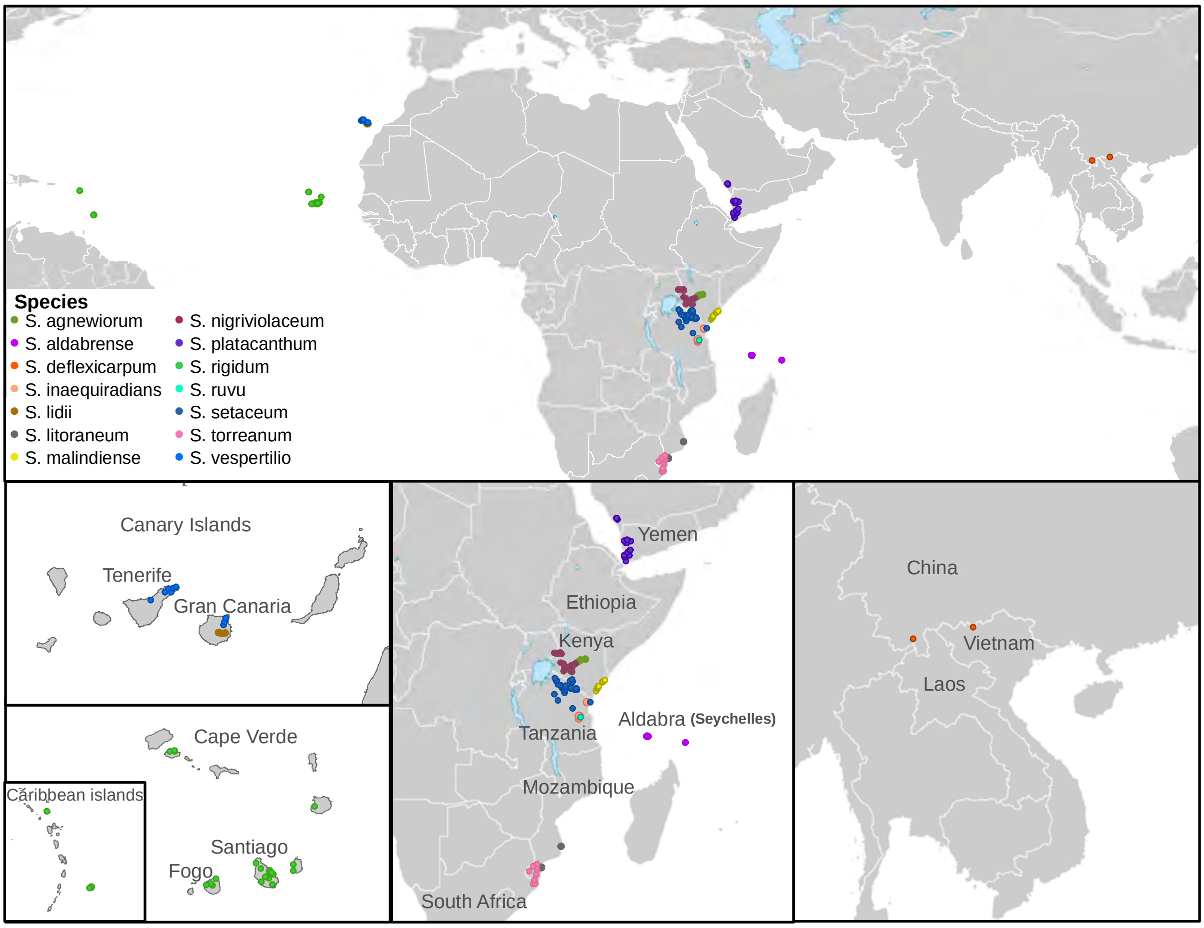


Figure 4.3: Map of georeferenced specimens of eggplant wild relatives identified as at risk of extinction. One species (*S. lidii*) has been formally assessed as critically endangered, nine species (*S. agnewiorum*, *S. aldabrense*, *S. deflexicarpum*, *S. inaequiradians*, *S. litoraneum*, *S. malindiense*, *S. setaceum*, *S. torreanum*, *S. vespertilio*) were identified in this study as threatened and three (*S. nigriviolaceum*, *S. platacanthum*, *S. rigidum*) as near threatened according to the preliminary IUCN Red List assessment criteria, and one (*S. ruvu*) has been previously suggested to be extinct in the wild.

occurring in one pixel) and also containing Red List species are protected (Figure 4.4b). More than half of these target collecting priorities occur in the hot semi-arid habitats generally known as the southern Acacia-Commiphora bushlands and thickets ecoregion (Olson *et al.*, 2001) and are the southern edge of White's (1983) Somali-Masai Acacia-Commiphora deciduous bushland and thicket phytochorion. More recent satellite images show that a significant proportion of habitats in this region have been converted to croplands (European Space Agency, 2009) and that vegetation today is a complex mosaic.

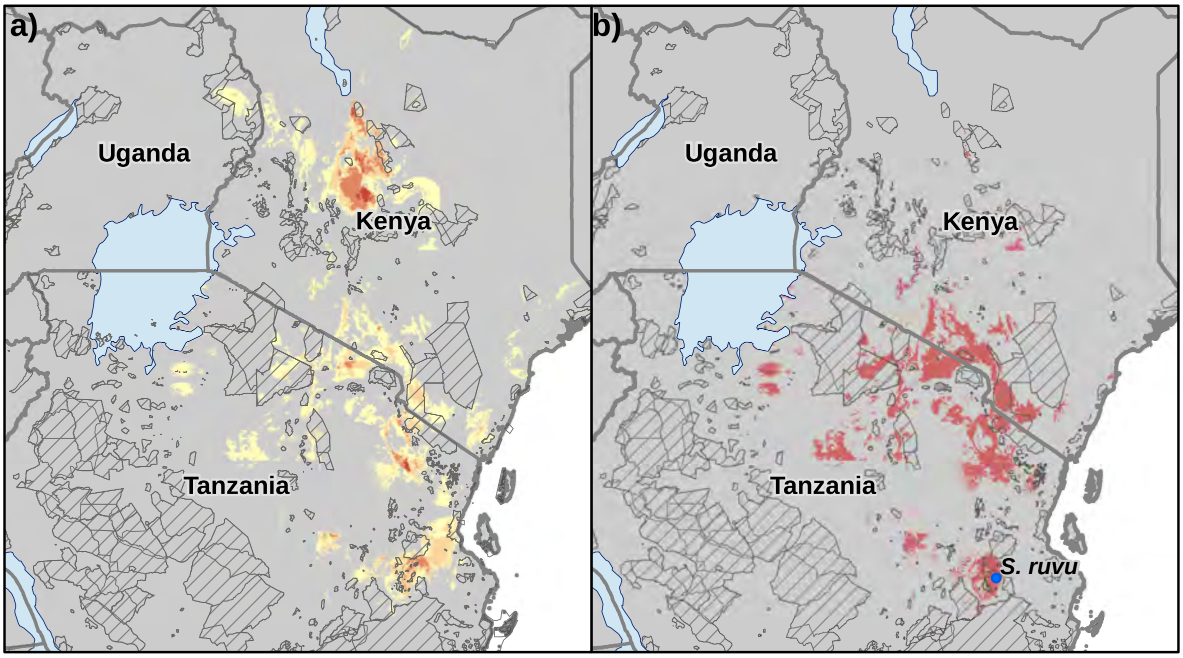


Figure 4.4: Hotspots (7 or more species per pixel) in relation to protected areas in eastern Africa. Hatched polygons are Protected Areas (World Database on Protected Areas Consortium, 2010). The georeferenced type specimen for *S. ruvu* (possibly extinct in the wild) is included. Further collecting priority hotspots: a) with all species of eggplant wild relatives; b) with at least one species that has a preliminarily threatened status on the Red List.

4.5 Discussion

Defining gene pools in large genera such as *Solanum* is challenging and we have approached this with a compromise solution that takes into account phylogenetic relationships and crossability (Vorontsova *et al.*, 2013; Plazas *et al.*, 2016). The concept presented here uses primarily phylogenetic relationships to define GP2, and GP3 includes less closely related but useful species which have previously been used in eggplant breeding but that thus far cannot be easily crossed with *S. melongena* (Plazas *et al.*, 2016). Our new suggested gene pool concept for eggplant includes one wild species (the wild progenitor of the eggplant, *S. insanum*) in GP1, 48 species in GP2, and three more widespread weedy species with native distributions in the New World in GP3 (see Table 4.2). Our revised gene pool based on phylogenetic and taxonomic criteria better represents the true wild diversity of eggplant CWR, and provides an initial framework for current and future breeding efforts. A more systematic approach to record results of trials of interspecific crosses between cultivated eggplant with its wild relatives can help in producing a refined gene pool concept more aligned with the Harlan and de Wet (1971) approach, that would be widely useful in the plant breeding community. Our results here focus on the wild species, but in eggplants and many other crops the dividing line between wild and cultivated can be difficult to determine (see Davidar *et al.*, 2015).

The use of a broad gene pool concept is potentially important to best capture traits desirable for crop improvement that may be present in species not closely related to or directly crossable with the crop. In this case, the selection of those distantly related, but potentially useful taxa might best look at weedy, invasive species, whose resistance to both biotic and abiotic stresses are likely to be well-developed (Daunay and Hazra, 2012). Our use of three New World species of spiny solanums that are distantly related to the eggplant is an illustration of this concept. We deliberately chose species that had been used in eggplant breeding, and that are also were widespread outside their putative native ranges. Widespread species may hold a wealth of genetic variation and adaptive traits and thus may be of great significance to breeders in the long run, especially in light of new techniques for hybridization and gene editing.

Despite the recent taxonomic revision (Knapp *et al.*, 2013) of species-level identities and names for members Eggplant clade, taxonomic challenges persist within wild eggplants. A systematic specimen identification of eggplant has been carried out and it is reflected in the occurrence records of herbarium samples derived from Solanaceae Source. Germplasm collections, however, clearly require re-identification, in part due to the revised species circumscriptions of Knapp *et al.* (2013) but also because this information is not frequently revised or updated. For instance, *S. campylacanthum* accessions are often named as *S. incanum* (groups A and B) in genebanks, but we found many genebanks did not include the letter designation that would allow us to directly compare identifications (see also Meyer *et al.* 2012a, where all close wild relatives were labelled as *S. incanum*). Consequently it is likely that germplasm collections are underestimated for *S. campylacanthum* and overestimated for *S. incanum*.

Overall, the gap analysis reveals significant gaps in genebank holdings of eggplant CWR with 40 out of 52 species ranked as medium or high priority for future collecting for *ex situ* conservation. Most of the species currently considered of high value for eggplant breeding

(Daunay and Hazra, 2012; Rotino *et al.*, 2014) are currently poorly represented in genebank collections. Improving the current representativeness of eggplant CWR in genebanks is vital for conserving the genetic diversity in these species, while at the same time widening source options for mining of valuable traits by eggplant breeders.

With large geographic regions identified as priorities for further collecting, our additional analyses to refine the collecting priorities based on IUCN Red List threat assessment categories can help to prioritize the most threatened species (see final column in Table 4.3). Geographic collecting gaps often coincide with areas known to have degraded habitats due to conversion to agriculture and urbanization (Vorontsova *et al.*, 2010), so geographic and species collecting gaps can also directly inform strategies for the management of threatened species *in situ*. Of all prioritized areas, the Tanzania-Kenya border is remarkable in possessing as many as 10 species considerably under-represented in genebanks, of which four are among the closest relatives of the cultivated eggplant and five are at potential risk of extinction.

Three important facets for future collecting of eggplant CWR for conservation are of importance in prioritization of germplasm acquisition. Firstly, those CWR species that are threatened in their native ranges, secondly those whose populations harbor significant genetic diversity, and thirdly those species whose biology and distribution suggests they might harbor traits of significance for eggplant breeding in the future.

Of the several eggplant wild relatives identified here as threatened with extinction, the only one formally listed as critically endangered on the Red List is the Canary Island endemic *S. lidii* (see Figure 4.3; Rodríguez Delgado *et al.*, 2011). The population size and genetic diversity of extant populations of *S. lidii* is limited, which may compromise their medium term viability (Prohens *et al.*, 2007). Fortunately 14 accessions of the species are conserved in genebanks (see Table 4.2), and wild populations occur in the Reserva Natural Especial de Los Marteles on the island of Gran Canaria (Rodríguez Delgado *et al.*, 2011), where it may be able to continue to evolve in its natural habitat. The design and implementation of management plans for its *in situ* conservation will be vital to secure the long-term conservation of this species. Conversely, both *S. deflexicarpum* and *S. agnewiorum* have small range sizes, occur in areas of high habitat degradation, but have no germplasm accessions in genebanks. *Solanum deflexicarpum* is a member of the Anguivi grade closely related to *S. violaceum* (Aubriot *et al.*, 2016), a species previously used in eggplant breeding (see Table 4.1) and is here assessed as endangered. With a very small range size and a fragmented distribution (see Figure 4.3) *S. deflexicarpum*, despite being less threatened as assessed using the IUCN criteria, is clearly important for future collecting and use in crop improvement. The same holds true for *S. agnewiorum*, only recently described from eastern Africa. This forest species is a surprising member of the Eggplant clade (Vorontsova *et al.*, 2013; Aubriot *et al.*, 2016) and is thus a prime candidate for crossing experiments with the cultivated eggplant.

Widespread species are also important reservoirs of genetic variation for potential use in crop improvement. *Solanum insanum*, the wild progenitor of eggplant, is extremely widely distributed from Madagascar throughout south and southeastern Asia (as far east as the Philippines). Meyer *et al.* (2012a) working the Asian range of both *S. melongena* and *S. insanum* (identified in their study as *S. incanum*) found considerable genetic diversity within both species that in part paralleled geography, but not completely. Mutegi *et al.* (2015) worked with populations of *S. insanum* in southern India and showed that populations of the wild species (*S. insanum*) were genetically distinct from the cultivated plants, and were highly structured geographically. They separated *S. insanum* into three major genetic groups and suggested that geographic isolation and limited gene flow have structured the genetics of this species in the study area. This high genetic divergence among populations suggests that further population level sampling across the range of *S. insanum* will uncover genetic novelty worth conserving in genebanks, especially given the lack of reproductive barriers between *S. insanum* and *S. melongena* (Davidar *et al.*, 2015; Plazas *et al.*, 2016).

Solanum campylacanthum is equally widespread in Africa, and occurs in a huge variety of mostly disturbed and human-altered habitats (Vorontsova and Knapp, in press). Leaf shape is particularly variable in this species (see Figure 2 in Knapp *et al.* 2013), leading to its having more than 70 synonyms and having been classified as *S. incanum* group A (wider leaves) and *S. incanum* group B (narrower leaves) in previous classification systems (Lester and Hasan, 1991). Morphological variability of this magnitude is likely to be mirrored in genetic variability, although this is not always the case (Kardos *et al.*, 2015). This again suggests that for conservation in genebanks, populations across the range of *S. campylacanthum* should be collected. Our analyses of hotpots of collecting gaps for all species identify an efficient way in which to begin such a challenging collecting program.

Weeds are plants that are successful at colonizing new areas and often are those found in disturbed areas. Many of the eggplant CWR analysed here could be classified as weeds: *S. campylacanthum*, *S. insanum*, and *S. anguivi* all have very large distributions in many different sorts of disturbed and semi-disturbed habitats (Vorontsova and Knapp, in press). Weediness facilitates commensalism with humans, ease of cultivation, and possibly also crossability with domesticates, but not necessarily other traits thought to be important for domestication such as loss of secondary metabolites or changes to fruit and seed morphology (see Meyer *et al.* 2012b for a discussion and critique of domestication syndrome traits). Weedy species are also often resistant to many pests and diseases; biotic and abiotic resistance may help to account for their broad ranges. The species of CWR that have to date provided useful pest and disease resistance have all been weedy (Table 4.1), but most of these are only partially interfertile with *S. melongena* (Daunay and Hazra, 2012) or require more advanced techniques (e.g., somatic hybrids and tissue culture) to interbreed (e.g., Daunay *et al.*, 1993; Kumchai *et al.*, 2013) or, alternatively, are used as rootstocks (Rahman *et al.*, 2002; Prohens *et al.*, 2013). Richardson and Rejmánek (2011) included the eggplant CWR *S. torvum* and *S. sisymbriifolium* in a recent global analysis of invasive species, and it could be argued that invasives are the ultimate weeds. We included three distantly related, weedy (possibly invasive) species in our GP3 for eggplant and suggest that in including such taxa in a gap analysis for germplasm collecting both native and non-native ranges should be part of the analysis. Populations of weedy (invasive) species outside their native range may have escaped from local pests and diseases and exhibiting ecological release, or conversely they could be very resistant to whatever biotic challenges they confront. Distinguishing between these two possibilities is important in the context of collecting CWR for conservation in genebanks, especially if weedy CWR are resistant to a great variety of challenges. In this context *S. viarum* Dunal and *S. torvum* (both included in our GP3) may be particularly interesting because their non-native ranges include regions in southeastern Asia where eggplant is thought to have been domesticated (see Figure 4.1) and presumably also where pests and disease that have co-evolved with *S. melongena* occur.

Our study using herbarium specimen data identified and georeferenced to a high standard and compared to data taken from genebank collections has shown that eggplant wild relatives are extremely poorly represented, both taxonomically and geographically, in genebanks worldwide. Our high resolution species distribution models can be of use in targeting areas for further collecting in order to improve the conservation status of these species *ex situ*. Combining the *ex situ* gap analysis with an assessment of conservation threat *in situ* has revealed hotspots for conservation action; many of the species under the most threat exist only outside of current protected areas. Collection of these species will also be important for biodiversity conservation *ex situ*. Earlier unpublished versions of our analyses have already been useful to prioritize collecting efforts of eggplant CWR species in eastern Africa (Dempewolf *et al.*, 2013). Working with traditional gene pool concepts is challenging in large genera such as *Solanum*, but using a combined approach and incorporating recent taxonomic and phylogenetic advances we have identified both species and regions for further germplasm collecting to both increase understanding of eggplant wild relatives and strengthen eggplant breeding in the future.

CHAPTER 5

GLOBAL CONSERVATION PRIORITIES FOR CROP WILD RELATIVES

Article accepted for publication by Nature Plants. DOI: `http://dx.doi.org/10.1038/NPLANTS.2016.22`

Nora P. Castañeda-Álvarez*, Colin K. Khoury*, Harold A. Achicanoy, Vivian Bernau, Hannes Dempewolf, Ruth J. Eastwood, Luigi Guarino, Ruth H. Harker, Andy Jarvis, Nigel Maxted, Jonas V. Müller, Julián Ramírez-Villegas, Chrystian C. Sosa, Paul C. Struik, Holly Vincent and Jane Toll.

N.P.C.-A., C.K.K., H.D., R.J.E., L.G., A.J., N.M., J.M., J.R-V. and J.T. conceived and designed the study. N.P.C.-A., C.K.K., H.D., R.J.E., R.H.H., A.J., N.M., J.R-V., C.C.S. and H.V. acquired and contributed data. N.P.C.-A., C.K.K., H.A.A., V.B. and C.C.S. processed the data, performed the analyses and analyzed the results. N.P.C.-A., C.K.K., H.D., R.J.E., L.G., A.J., N.M. and J.M. interpreted the results and wrote the manuscript. N.P.C.-A., C.K.K., V.B., H.D., R.J.E., L.G., A.J., N.M., J.M., J.R-V. and P.C.S. edited the manuscript.

* These authors contributed equally to this work. Details of our contributions are given in the Declaration of this thesis.

5.1 Abstract

The wild relatives of domesticated crops possess genetic diversity useful for developing more productive, nutritious, and resilient crop varieties. Their conservation status and availability for utilization are a concern and have not been quantified globally. Here we model the distributions of 1,076 taxa related to 81 crops of primary importance to food security. We analyze the extent to which the potential diversity encompassed in these taxa is currently accessible to the global research community through genebanks. Our results indicate that the diversity of crop wild relatives is currently poorly represented in genebanks, and a systematic effort is needed to improve their conservation and availability for use in plant breeding. For 313 (29.1% of total) taxa associated with 63 crops, no germplasm accessions exist, and a further 257 (23.9%) are represented by fewer than 10 accessions. Over 70% of taxa are identified as high priority for further collecting in order to improve their representation in genebanks, and over 95% are insufficiently represented in regard to the full range of geographic and ecological variation in their native distributions. The most critical collecting gaps occur in the Mediterranean and Near East; western and southern Europe; Southeast and East Asia; and South America.

5.2 Introduction

The challenges to global food security are complex and compounding. Our growing population and changing dietary expectations are projected to increase demand on food systems for at least the next four decades (Godfray *et al.*, 2010; Foley *et al.*, 2011; Tilman *et al.*, 2011; Tilman and Clark, 2014; Khoury *et al.*, 2014), outpacing forecasted crop yield gains (Ray *et al.*, 2013). Limitations in land, water, and other natural resource inputs, competition for arable soils with non-food crops and other land uses, soil degradation, climate change, and the need to minimize harmful impacts on ecosystem services and biodiversity further constrain production potential (Cordell *et al.*, 2009; Godfray *et al.*, 2010; Foley *et al.*, 2011; Asseng *et al.*, 2015). Although gains in food availability may partially be obtained through dietary change and food waste reduction (Foley *et al.*, 2011; Tilman and Clark, 2014), increases in the productivity, resilience and sustainability of current agricultural systems are clearly necessary (Tilman *et al.*, 2011). Key to this sustainable intensification is the use of novel genetic diversity in plant breeding to produce crop varieties containing traits such as drought and heat tolerance, increased pest and disease resistance, and input use efficiency (Zhu *et al.*, 2000; Esquinas-Alcázar, 2005; McCouch *et al.*, 2013).

As sources of new genetic diversity, crop wild relatives - the wild cousins of cultivated plant species - have been used for many decades for plant breeding, contributing a wide range of beneficial agronomic and nutritional traits (Xiao *et al.*, 1996; Tanksley and McCouch, 1997; Gur and Zamir, 2004; Hajjar and Hodgkin, 2007; McCouch *et al.*, 2007; Guarino and Lobell, 2011). Their utilization is expected only to increase as a result of ongoing improvements in information on species and their diversity and advances in breeding tools (Tanksley and McCouch, 1997; McCouch *et al.*, 2012). However, this expectation is based on the assumption that crop wild relatives will be readily available for research and plant breeding, which requires their conservation as germplasm accessions in genebanks, as well as functioning mechanisms to enable access to this diversity (Esquinas-Alcázar, 2005; McCouch *et al.*, 2013). Preliminary assessments of the comprehensiveness of conservation of wild relatives in genebanks have suggested substantial gaps (FAO, 2010; Vincent *et al.*, 2013), and wild populations of a range of species are threatened by the conversion of natural habitats to agriculture, urbanization, invasive species, mining, climate change and/or pollution (Wilkes, 2007; Jarvis *et al.*, 2008; Brummitt *et al.*, 2015b). A concerted effort devoted to improving the conservation and availability of crop wild relatives for crop improvement is thus timely both for biodiversity conservation and food security objectives (Dempewolf *et al.*, 2013), as the window of opportunity to resolve these deficiencies will not remain open indefinitely (Wilkes, 2007; FAO, 2010).

We conducted a detailed analysis of the extent of representation of the wild relatives of 81 crops in genebanks equipped to provide access to these genetic resources to the global research and breeding community. The crops include major and minor cereals, root and tuber crops, oilcrops, vegetables, fruits, forages, and spices, chosen based on their importance to food security, income generation, and sustainable agricultural production (Supplementary Table 5.1). We first modelled the geographic distributions of a total of 1,076 unique crop wild relative taxa from 76 genera and 24 plant families (Supplementary Table (Supplementary Table 5.2). We then compared the potential geographic and ecological diversity encompassed in these distributions to that which is currently accessible in genebanks (Ramírez-Villegas *et al.*, 2010). In order to aid conservation strategies, we categorized taxa with a “final priority score” (FPS) for further collecting from the natural habitats of crop wild relatives to increase representation in genebanks, on a scale from zero to ten. The final priority score was created by averaging each taxon’s assessed current representation in genebanks in regard to overall number of accessions, geographic diversity, and ecological diversity. High priority for further collecting was assigned for taxa where $\text{FPS} \geq 7$ (i.e., very little or no current representation in genebanks); medium priority where $5 \leq \text{FPS} < 7$; low priority where $2.5 \leq \text{FPS} < 5$; and sufficiently represented for taxa whose $\text{FPS} < 2.5$. Finally, we identified geographic hotspots where considerable richness of high priority wild relative taxa is concentrated. Such sites represent particularly valuable targets, both for efficient collecting for *ex situ* conservation in genebanks as well as for *in situ* conservation in protected areas.

5.3 Methods

Methods used for gathering data, modelling, analyses, and the associated references are available in the Supplementary File 5.1.

5.4 Results

The distributions of crop wild relatives were modelled to occur on all continents except Antarctica, and throughout most of the tropics, subtropics, and temperate regions, except the most arid areas and polar zones (Fig. 5.1). The greatest richness of taxa was modelled in the Mediterranean, Near East, and southern Europe, South America, Southeast and East Asia, and

Mesoamerica, with up to 84 taxa overlapping in a single 25 km^2 grid cell. These richness hotspots largely align with traditionally recognized centres of crop diversity (Vavilov, 1926), although the analysis also identified a number of less well-recognized areas, e.g., central and western Europe, the eastern USA, southeastern Africa, and northern Australia, which also contain considerable richness. Hotspots in tropical and subtropical areas also largely aligned with zones recorded as possessing high richness of endemic flora and fauna, and experiencing exceptional degrees of loss of habitat (Myers *et al.*, 2000). Temperate regions identified under the same criteria, e.g., the California and Cape Floristic Provinces, southwestern Australia, central Chile, and New Zealand, had considerably less overlap with areas rich in crop wild relatives.

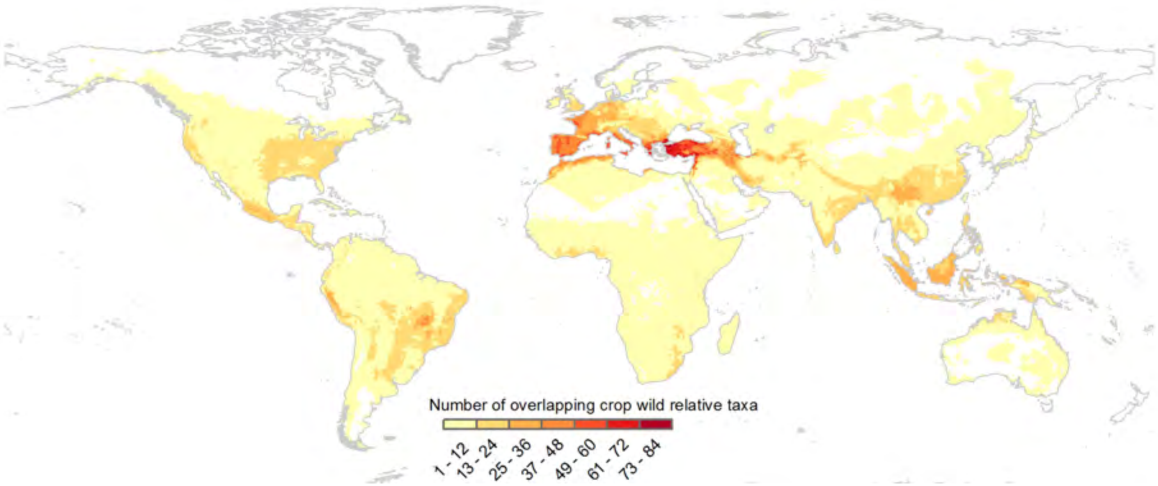


Figure 5.1: Crop wild relative taxon richness map. Map displays overlapping potential distribution models for assessed crop wild relatives. Darker red colours indicate greater overlap of potential distributions of taxa, i.e., where greater numbers of crop wild relative taxa occur in the same geographic area.

Wild relative taxa as a class of plant genetic resources were found to be critically underrepresented in genebanks. For 313 (29.1% of total) taxa associated with 63 crops, no germplasm accessions exist at all, and a further 257 taxa are represented by fewer than ten accessions. A total of 765 (71.1%) taxa were ranked as high priority for further collecting from their natural habitats, 148 (13.8%) as medium priority, 118 (11.0%) as low priority, and only 45 (4.2%) as currently sufficiently represented in genebanks (Supplementary Table 5.2). The mean FPS across all species [7.9 ± 2.5 (mean ± SD)] fitted well within the high priority category range. Lack of geographic and ecological representation in genebanks contributed significantly to most of the high FPS values, whereas less extreme gaps were generally evident in the total numbers of accessions conserved (Supplementary Fig. 5.1).

An analysis of wild relatives grouped by their associated crop (i.e., by crop gene pool) revealed that 72% of the crop gene pools were assigned high priority for further collecting (as an average of FPS scores across associated wild relative taxa), and thus require urgent conservation action (Figs 5.1 and 5.2). These included the gene pools of commodity crops of critical importance to global food supplies and/or agricultural production, e.g., sugarcane (9.2 ± 1.6), sugar beet (8.1 ± 1.6), and maize (6.9 ± 2.1), as well as important food security staples such as banana and plantain (9.4 ± 0.8), cassava (9.0 ± 1.6), sorghum (8.8 ± 1.0), yams (8.5 ± 2.9), cowpea (8.4 ± 1.7), sweetpotato (8.4 ± 1.7), pigeonpea (8.4 ± 1.1), millets (8.4 ± 2.7) and groundnut (7.6 ± 1.8) (Fig. 5.2 and Supplementary Table 5.1). High priority was also assigned to the gene pools of numerous crops important for smallholder income generation in the tropics (e.g., cacao and papaya), minor crops increasing in popularity due to their nutritional qualities (e.g., quinoa), as well as various other important fruits (e.g., grape, apple, watermelon, orange and mango), oilcrops (rapeseed), and forages (alfalfa) possessing considerable numbers of wild related taxa. Although all gene pools contained taxa with considerable conservation concerns, the wild relatives of fruits, forages, sugar crops, starchy roots, and vegetables were those assessed as least well represented in genebanks (Supplementary Fig. 5.2). Average FPS values across all wild relatives per crop type were 8.8 ± 1.8 for fruits, 8.7 ± 1.7 for forages, 8.6 ± 1.6 for sugar crops, 8.2 ± 2.3 for starchy roots, 8.1 ± 2.4 for vegetables, 7.1 ± 2.6 for pulses, 7.1 ± 2.3 for oilcrops, 7.1 ± 1.9 for spices, and 6.4 ± 3.1 for cereals.

None of the 81 assessed crop gene pools demonstrated an average FPS across its wild relatives that would permit its categorization as sufficiently well represented in genebanks (Fig. 5.3). The wild relatives of six crops were assessed as fairly well represented, i.e., low current priority for further collecting for the gene pools of wheat (3.7 ± 2.4), grasspea (3.7 ± 2.0), chickpea (4.2 ± 2.6), and tomato (4.5 ± 1.9). Wheat and tomato, along with medium priority crop gene pools such as sunflower (6.3 ± 2.2), rice (6.6 ± 2.5), and potato (6.7 ± 2.6) have a long history of use of wild relatives in crop improvement (Xiao *et al.*, 1996; Zhu *et al.*, 2000)

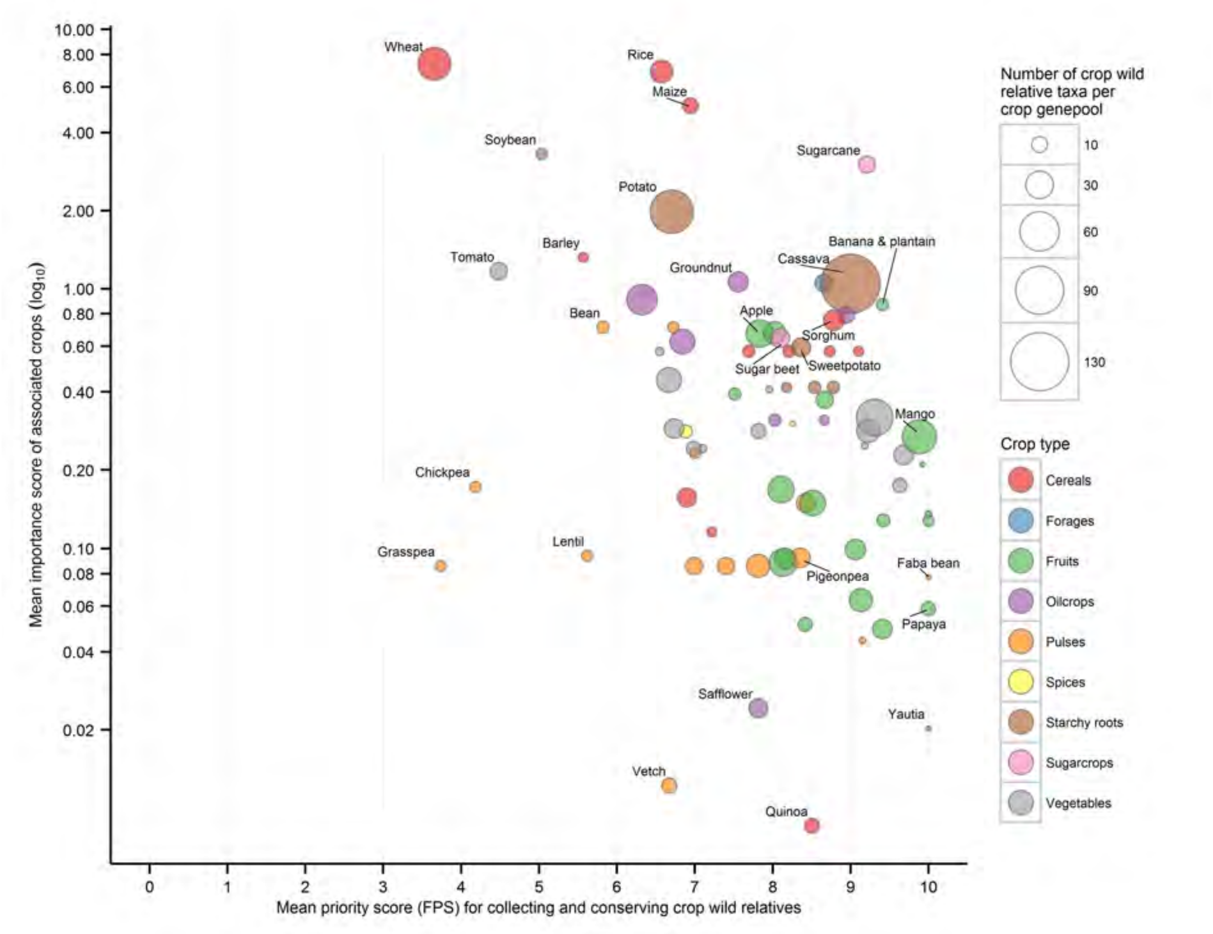


Figure 5.2: Collecting priorities for crop wild relatives and importance of associated crops. The priority scale displays the average of Final Priority Scores (FPS) across wild relatives per crop. The mean importance class of associated crops displays the significance of crops averaged across global food supplies and agricultural production metrics (see Supplementary File 5.1). For both axes, the scale is zero to ten, with ten representing the highest priority for further collecting/most important crop. The size of crop gene pool circles denotes the number of wild relative taxa per crop, ranging from 1 (faba bean) to 135 (cassava).

and benefit from relatively extensive germplasm collections. Other crop gene pools determined as low priority (e.g., grasspea and chickpea) have few wild relatives, and these generally present restricted distributions that have been fairly well sampled. However, specific taxa were assessed as under-represented in genebanks even within these low priority gene pools. For example, five taxa related to wheat were assessed as medium or high priority, one taxon related to grasspea as medium priority, three taxa related to chickpea as medium priority, and six taxa related to tomato as medium or high priority (Supplementary Table 5.2).

Proposed hotspots for further collecting for high priority crop wild relatives were identified

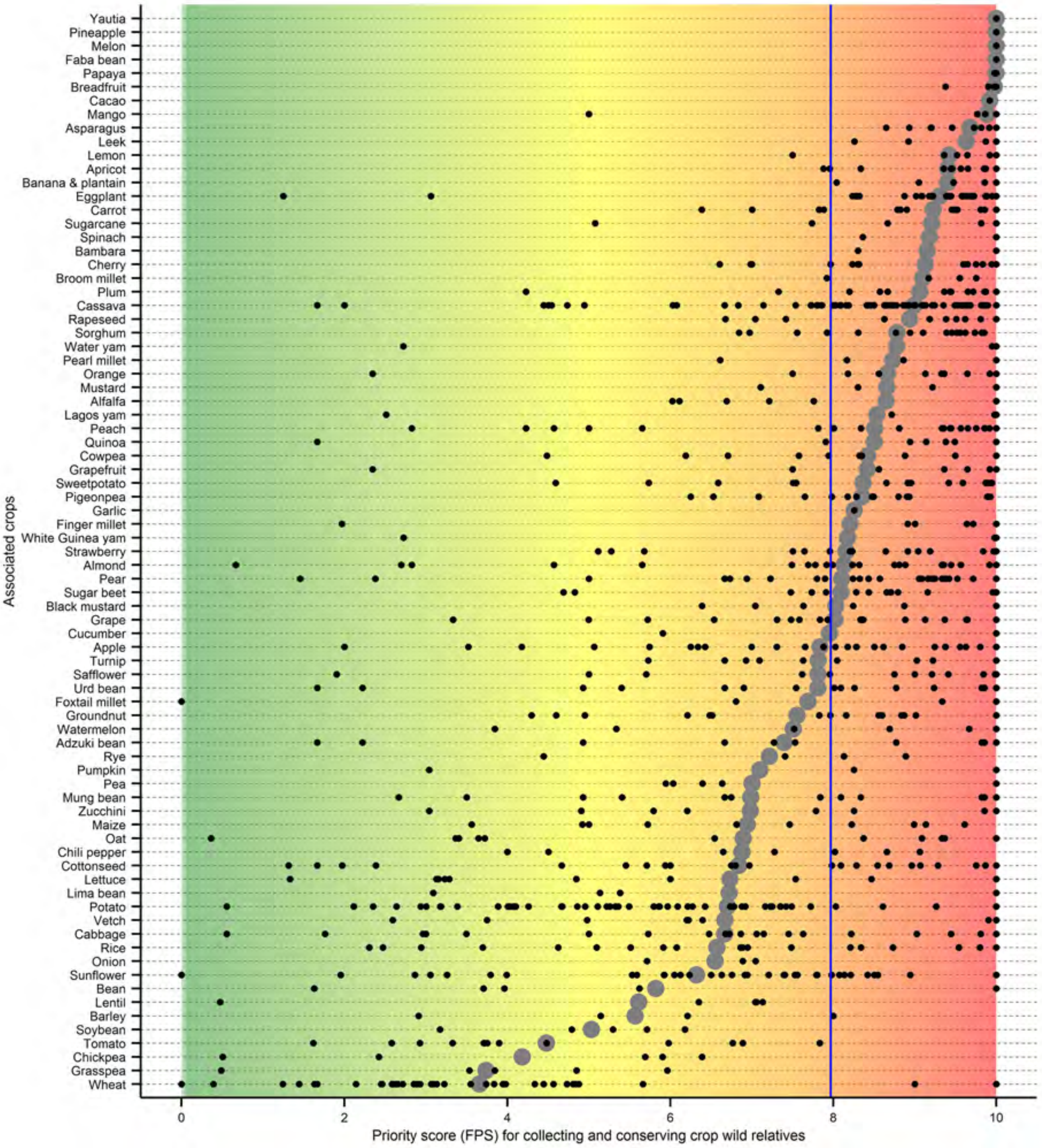


Figure 5.3: Collecting and conservation priorities for crop wild relatives by associated crop. Black dots represent the Final Priority Scores (FPS) for further collecting for wild relative taxa, with gray larger dots representing the average FPS across taxa per crop gene pool. The blue straight vertical line represents the mean FPS across all crop wild relative taxa within all crop gene pools.

across the world's tropical, subtropical, and temperate regions, with the most critical gaps identified in the Mediterranean, Near East, and southern and western Europe; Southeast and East Asia; and South America (Fig. 5.4). Up to 43 wild relative taxa (main map in Fig. 5.4) associated with up to 23 crops (inset map in Fig. 5.4) may potentially be collected within a single 25 km^2 grid cell.

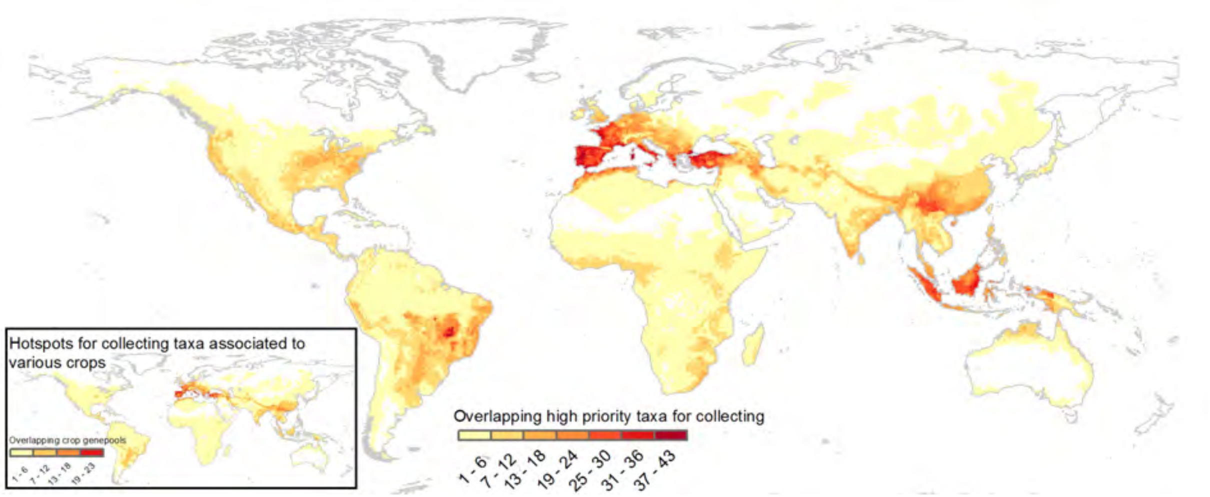


Figure 5.4: Proposed hotspots for further collecting activities for high priority crop wild relatives. The map displays geographic regions where high priority crop wild relative taxa are expected to occur and have not yet been collected and conserved in genebanks. The inset map shows gaps for under-represented taxa by crop gene pool. Darker red colours indicate greater overlap of potential distributions of under-represented taxa, i.e., where greater numbers of underrepresented crop wild relative taxa occur in the same geographic area. For the inset map, greater numbers indicate greater overlap of taxa associated to various crops.

5.5 Discussion

Our results demonstrate that crop wild relatives are currently under-represented and a systematic effort to improve their comprehensiveness in genebanks is critically needed. These findings are remarkable given the extensive efforts particularly in the past half century by international, regional, and national initiatives to conserve the broad diversity of important agricultural crops (Esquinas-Alcázar, 2005; FAO, 2010). Achieving the comprehensive conservation of crop genetic resources *ex situ* is constrained by technical as well as political and funding challenges, particularly over recent decades (Esquinas-Alcázar, 2005), and is most poignant for wild taxa, which are less well researched than crop species and often more difficult to conserve and to utilize (Esquinas-Alcázar, 2005; FAO, 2010; Dempewolf *et al.*, 2013). Addressing conservation gaps globally for crop wild relatives, a goal which is specifically targeted in recent major international agreements (i.e., The United Nations Sustainable Development Goals and the Strategic Plan for Biodiversity (Secretariat of the CBD, 2010)) will require substantial investment and extensive international collaboration. The high spatial resolution of these results is already informing such initiatives (Dempewolf *et al.*, 2013) and can be useful to the development of further efforts.

Here we outline priorities for collecting wild relatives based upon their current representation in genebanks, along with an assessment of the relative importance to global food supplies and production systems worldwide of their associated crops (Fig. 5.2; Supplementary Fig. 5.2). We recommend filling gaps in *ex situ* conservation first for the crop wild relatives of significant importance to global and regional food security which are sparsely represented in genebanks, e.g., maize, cassava, bananas and plantains, sorghum, millets, sweetpotato, yams, groundnut, cowpea, and pigeonpea.

In order to further refine these priorities, additional information and filters are needed. These include incorporating knowledge of threats to populations due to habitat modification, climate change, and other impacts. Preliminary field surveys and threat analyses for under-represented taxa are therefore urgently needed. We note that extensive expert evaluations of the results generally confirmed the robustness of our species distribution models and conservation prioritizations but also clearly emphasized the need to address urgent threats to the survival of many crop wild relative populations (Supplementary Fig. 5.3). Realistic strategies for field collection and subsequent *ex situ* conservation resulting in an increased availability of germplasm for plant breeding also require negotiating policy governing germplasm collecting and exchange (FAO, 2010; Secretariat of the CBD, 2011), assessing field work risks (e.g., war and civil strife in regions with high levels of diversity of wild relatives), coordinating timing of field work to maximize the collection of viable seeds and other propagules, prioritizing target crop gene pools based upon the interest of the breeding community in utilizing wild germplasm, and determining the relative difficulty of maintenance of targeted wild germplasm in genebanks. While the seeds of most wild relatives can be maintained under standard conditions for long-term conservation *ex situ*, some wild relatives produce recalcitrant seeds or do not produce seeds at all.

Such wild relatives may require more expensive approaches (e.g., *in vitro* or cryopreservation), and particularly for such taxa alternative conservation strategies such as the establishment of *in situ* conservation reserves may be more effective.

Despite an extensive effort to compile occurrence records from more than 400 different data sources, the wild relatives of a number of important agricultural crops (namely coffee, tea, and avocado) were not assessed due to lack of sufficient accessible data. We also note that a number of agricultural crops are not currently known to possess closely related wild relatives, including taro (*Colocasia esculenta*), coconut (*Cocos nucifera*), and date palm (*Phoenix dactylifera*). Improvements in the generation and accessibility of taxonomic, relatedness and geographic information on wild relatives (Wiersema *et al.*, 2012; Vincent *et al.*, 2013) may permit conservation assessments for some of these gene pools in the future.

The combination of the sampling, geographic and ecological representativeness scores used to determine the extent of conservation of the wild relatives of important agricultural crops in genebanks represents an efficient methodology for prioritizing taxa across crop gene pools given wide variations in the potential diversity encompassed in each taxon and the general absence of molecular data for such species. The sampling representativeness score permitted an indication of the total number of germplasm accessions estimated as sufficient to represent a taxon, relative to the known extent of the taxon and utilizing all genebank and reference data regardless of whether geographical coordinates are available. The geographic and ecological variation metrics were used as proxy for genetic diversity and potential functional adaptation to diverse environments, based on the assumption that the genetic composition of plant species varies across geographic range and is associated with adaptation to different ecological conditions (Eckert *et al.*, 2008). The increasing power and decreasing costs of direct measures of diversity in genomes may make significant future refinements of priorities achievable (McCouch *et al.*, 2013). However, further collecting is still needed for a very large number of wild relatives in order to assemble sufficient samples to perform such genetic assessments, as well as to help resolve taxonomic and gene pool assignment uncertainties (Goodwin *et al.*, 2015).

CHAPTER 6

CLIMATE CHANGE IMPACTS ON THE DISTRIBUTIONS OF CROP WILD RELATIVES

6.1 Summary

Here I analyze the potential effects of climate change on the distributions of the wild relative taxa of 29 crops important for global food security and sustainable agriculture. In this analysis, I consider two different adaptation scenarios, one assuming that wild relative taxa are not able to track their suitable climates, and the second one assuming that wild relative taxa are able to track the suitable climates that ensure their survival. This analysis uses an ensemble of 30 climate change projections for the RCP 4.5 scenario which stabilizes radiative forcing by the end of the century (2100) as a result of global actions to restrict greenhouse gases emissions. Results suggest that wild relative taxa analyzed are projected to lose in average 20.8% of their suitable areas. Only 28 taxa are under the risk of extinction as more than 50% of their areas are projected to change. I also discuss the implications for *in situ* and *ex situ* conservation under two adaptation scenarios.

6.2 Introduction

Life in all its forms is being affected by the increasing temperatures and changes in rainfall patterns that result from climate change (Settele *et al.*, 2014). In a global analysis of the impacts of climate change on more than 3,000 different species, Thomas *et al.* (2004) estimated that 15 - 37% of the species analyzed are likely to lose all their distribution area by 2050, and therefore are under high risk of being extinct. Changes in the composition of ecosystems and in the phylogenetic diversity contained within a region have been also projected as some of the potential impacts of climate change on animals and plants (Peterson *et al.*, 2002; Thuiller *et al.*, 2011).

Plants are one of the groups projected to be relatively more affected by climate change because of their sessile nature. Significant reductions in the geographical ranges of plant species have been projected, even under stringent mitigation scenarios (Warren *et al.*, 2013). Additionally, vital processes that secure the growth, development and reproduction of plants can also be disrupted due to warmer temperatures as it has been evidenced with the changes in the flowering patterns of *Rhododendron* species in the Himalayas (Hart *et al.*, 2014), in the pollination rates of orchids (Robbirt *et al.*, 2014), and germination and seedling establishment in alpine plants (Mondoni *et al.*, 2012). This panorama is aggravated by the current exposure of plants to additional factors that undermine their survival, such as the conversion of natural habitats for agriculture and livestock production, overharvesting, invasive species, and high fragmentation of natural habitats (Bilz *et al.*, 2011; Watson *et al.*, 2013; Brummitt *et al.*, 2015b).

Mechanisms such as phenotypic plasticity, adaptive evolution and dispersal may help plants to respond to climate change (Chen *et al.*, 2011; Franks *et al.*, 2014). However, the effectiveness of these mechanisms will depend on many factors such as the speed of global warming, the genetic composition of plant populations, the health of natural habitats, and the presence of competitors and pollinators (Chen *et al.*, 2011; Franks *et al.*, 2014; Alexander *et al.*, 2015).

As many other plants, crop wild relatives, a group of plants valued for their contributions of genetic material valuable for agriculture, are also exposed to the impacts of climate change. Jarvis *et al.* (2008) estimated that 16 to 22% of the wild relatives species of potato, cowpea and peanut may be under the risk of extinction by year 2050. Using species distribution modeling techniques, Ureta *et al.* (2012) estimates losses between 16 to 39% of the potential distributional ranges of the races and wild relatives of maize by 2050. Globally important crops have benefited from the use of wild relatives to produce more productive varieties (Gororo *et al.*, 2002), with improved nutritional content (Traka *et al.*, 2013) and tolerant to biotic and abiotic stresses (Rick and Chetelat, 1995; Ishimaru *et al.*, 2010; Jones *et al.*, 2014) (see Maxted and Kell (2009) for a review). Estimates of the monetary value of crop wild relatives in plant breeding are between USD 120 and 1,600 billion per year (Pimentel *et al.*, 1997; PWC, 2013; Tyack and Dempewolf, 2015).

Given the imminent threats to and evident benefits derived from plant biodiversity, policy frameworks to support the conservation and availability of biodiversity and genetic resources for future generations have been set up. The Convention on Biological Diversity on its Strategic Plan for Biodiversity 2011-2020 encourages participant countries to improve the current status of biodiversity and to safeguard species and genetic diversity as stated in its Strategic Goal C (Secretariat of the CBD, 2010). More recently, the United Nations' Sustainable Development Goals (SDGs) include an explicit call to stop biodiversity loss and to promote the access to genetic resources in its Goal 15. On the other hand, the International Treaty for Plant Genetic Resources for Food and Agriculture acts as a mechanism to promote the conservation and sustainable use of the plant genetic resources (including the crop wild relatives) of a subset of globally important crops (FAO, 2009).

Halting the loss of crop wild relatives and securing their future availability requires identifying the taxa considered as wild relatives of crops at the global (Vincent *et al.*, 2013) and national scales (Magos Brehm *et al.*, 2008; Khoury *et al.*, 2013; Fielder *et al.*, 2015); understanding their *in situ* and *ex situ* conservation status, and the threats that can affect their survival in the short and long term. This information can in turn help guiding conservationists, practitioners, governments and international organizations to set priorities and actions towards the conservation of these functional elements of biodiversity.

Despite the fact of knowing the projected negative effects of climate change on the distributions of plant species, including some crop wild relatives, extensive analysis for several wild relative taxa accross multiple crops have not been conducted. Here I present an analysis of the potential impacts of climate change on the distribution of the wild relatives of 29 crops. Towards this aim, I used environmental niche modeling techniques to estimate the potential geographic areas where the wild relative taxa are likely to occur currently and in the coming decades given the projected changes in climate, and two dispersal scenarios. I also discuss the implications of projected changes for the conservation of crop wild relative taxa, including possible adaptation strategies that may be required to address the potential climate change impacts.

6.3 Methodology

6.3.1 Crops and species selection

All food crops listed in the Annex I of the International Treaty on Plant Genetic Resources for Food and Agriculture (FAO, 2009), an established mechanism that facilitates access and promotes the use and conservation of plant genetic resources globally, were selected for this analysis. Once the crops chosen for the analysis were defined, the closest wild relatives associated to each crop were selected according to the degree of relatedness with their corresponding associated crop. The degree of relatedness provides information on the easiness to cross wild relative taxa with cultivated forms and the ability to produce fertile offspring. This relationship is described by the gene pool concept of Harlan and de Wet (1971), and surrogate classification schemes such as the Taxon Group concept Maxted *et al.* (2006), the Genetic-relative class Wiersema *et al.* (2012), and the provisional gene pool concept Vincent *et al.* (2013).

For the purpose of this analysis, the closest wild relative taxa associated to the crops previously selected were chosen as follows: Taxa in the primary and secondary gene pool concept following Harlan and de Wet (1971), taxa in the primary and secondary genetic-relative class according to Wiersema *et al.* (2012), taxa in the Taxon Groups 1a, 2, and 3 (Maxted *et al.*, 2006), and taxa in the primary and secondary provisional gene pools (Vincent *et al.*, 2013). Less related taxa were also included if they had a proven or potential use in breeding as reported in the literature. The degree of relatedness of wild relatives and their associated crops were retrieved from “The Harlan and De Wet Crop Wild Relative Inventory” (`www.cwrdiversity.`

`org/checklist/`) and “GRIN Taxonomy for Plants” (`www.ars-grin.gov/~sbmljw/cgi-bin/taxcrop.pl?language=en`) (Vincent *et al.*, 2013; USDA ARS National Genetic Resources Program, n.d.).

6.3.2 Occurrence data

Occurrence records of crop wild relative taxa were acquired, standardized and curated before being used as inputs of the environmental niche modelling algorithm. Existing digital databases were used as the main sources of occurrence records: The Global Biodiversity Information Facility (`www.gbif.org`), an international initiative that mobilizes data on all types of living creatures on Earth; The System-Wide Information Network for Genetic Resources (SINGER), the former system that centralized the passport information of germplasm accessions held within the CGIAR centers and the World Vegetable Center (AVRDC); and the Genetic Resources Information Network (GRIN), the US genebanks system. Other relevant sources of occurrence records include digital databases from individual institutions, national genebanks, researchers, previous projects, published and gray literature, and personal visits to herbaria.

Information recorded in botanical specimens, which were photographed during the visits to herbaria, were manually digitized and stored in an electronic spreadsheet, recording information on the specimen’s identity, geographical locality where the sample was taken (e.g., country name), geographical coordinates if available, description of the surrounding habitats, name(s) of collector(s), phenological traits visible in the specimen (i.e., availability of flowers, fruits or seeds), institutions where duplicates had been sent, type of specimen (e.g., isotype, haplotype) (Castañeda-Álvarez *et al.*, 2011). Once all individual datasets were digitized, their data fields were mapped to a larger database as a way to determine the type of information stored on each datafield, and in addition, those records originally provided without any geographical coordinate from the source were submitted to a batch-georeferencing process using GEOLocate (Rios and Bart, 2010). Only coordinates with a precision distance below 5km assessed in the subsequent steps.

Each individual dataset was standardized and curated before being merged into a single, larger database. For this, a two-step validation process was performed, one to verify the botanical nomenclature of each record, and the other one to assess the quality of geographic coordinates. The botanical identity validation consisted in verifying that the scientific name associated to each occurrence record did not contain misspellings and is aligned to accepted plant species checklists. Queries to GRIN (USDA ARS National Genetic Resources Program, n.d.), the Taxonomic Name Resolution Service (TNRS; Boyle *et al.*, 2013), and TaxonStand (Cayuela *et al.*, 2012) were performed, and the names retrieved were compared to the scientific name originally provided with each occurrence record. Scientific names in agreement with GRIN were assumed as accepted names. In the case when the scientific name did not coincide with GRIN, but it did coincide with the name suggested by TNRS or TaxonStand, TNRS was preferred over TaxonStand. For isolated cases, when the initial scientific name was not aligned with any of the sources retrieved, and additional revision was performed using JSTOR Global Plants (`plants.jstor.org`), the Encyclopedia of Life (Parr *et al.*, 2014), the International Plant Names Index (IPNI; *The International Plant Names Index*, 2012), and the Integrated Taxonomic Information System (ITIS; *Integrated Taxonomic Information System (ITIS)*, n.d.).

Once the botanical nomenclature of all data records was confirmed, the geographical coordinates assessment was performed. This assessment consisted in two steps: 1) Detecting off-land coordinates by mapping occurrence records over a global raster ($\sim$ 1 km resolution) and retaining only the occurrence points in land (Warren *et al.*, 2013); and 2) Resolving conflicts between the country where the sample was taken or recorded, and the country where the geographical coordinate is located when mapped over a global administrative units map (Hijmans *et al.*, 1999). Records where the originally registered country name coincides with the country where the coordinate is mapped are retained and considered as inputs for preparing the environmental niche models.

6.3.3 Current and future climate data

For the current climate, or baseline, I used an initial set of variables composed by one topographic variable and 19 bioclimatic variables obtained from WorldClim (`www.worldclim.`

`org`; Hijmans *et al.*, 2005) at a spatial resolution of 2.5 arc minutes (4.6 $\times$ 4.6 km at the Equator). These bioclimatic variables were produced by averaging climatological data between 1950 and 2000 (Hijmans *et al.*, 2005). An analysis to identify and to correct collinearity between environmental variables was applied. Collinearity, or high linear correlation between explanatory variables, may affect the performance of environmental niche models (Dormann *et al.*, 2013). Towards this aim, I used the Variance Inflation Factor (VIF) as it is a multivariable approach to identify highly correlated variables and subsequently reduce the number of environmental predictors to be used in MaxEnt. Variables with a correlation threshold above 10 were excluded from the final set of model predictors.

I used an ensemble of thirty Global Circulation Models (GCMs) produced by the Coupled Model Intercomparison Project Phase 5 (CMIP5) and used in the Intergovernmental Panel on Climate Change - Fifth Assessment (IPCC, 2014) of the Representative Concentration Pathway (RCP) 4.5 for the period between 2040 and 2069 (2050s) (Figure 6.1, Table 6.1). RCP 4.5 represents the trajectory of a stabilization scenario in which the radiative forcing is stabilized at 4.5 W m^{-2} (approx. 600 ppm CO_2-equivalent) near 2100 (Thomson *et al.*, 2011). RCP 4.5 assumes that stringent global policies to decrease greenhouse gas emissions have been put in place such as the valuation of carbon in natural vegetation, the implementation of reforestation programs to help decreasing the use of croplands and grasslands, the enhancement of crop yields and changes in human diets (van Vuuren *et al.*, 2011). All future climate data were downscaled to a resolution of 2.5 arc minutes using the delta method of Ramirez-Villegas and Jarvis (2010).

6.3.4 Environmental niche modelling

MaxEnt, a machine learning algorithm for modelling species environmental niches was selected thanks to its ability to discriminate the environmental niche of species using presence-only data, its overall performance when compared to other modeling algorithms, and extended use in ecological and conservation analyses (Elith *et al.*, 2006; Phillips *et al.*, 2006). The minimum required data for MaxEnt to produce an environmental niche model includes occurrence records with explicit geographic coordinates of the species to model, and a set of environmental layers

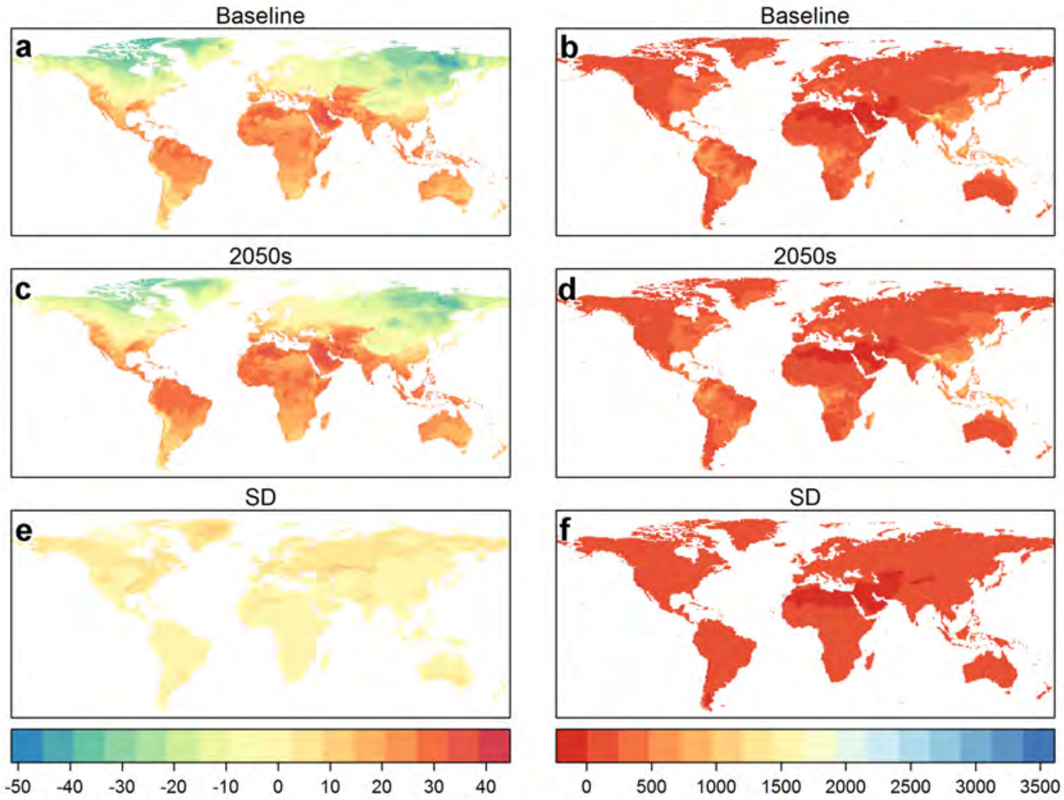


Figure 6.1: Current climatic conditions, future projected conditions and uncertainties. **a** Current mean temperature of driest quarter (BIO 9); **b** Current precipitation of warmest quarter (BIO 18); **c** Future mean temperature of driest quarter (BIO 9); **d** Future precipitation of warmest quarter (BIO 18); **e** Standard deviation of GCM projections of the mean temperature of driest quarter (BIO 9); **f** Standard deviation of GCM projections of precipitation of warmest quarter (BIO 18).

that the programme will use as predictor variables. Environmental niche models were only produced for taxa with more than ten unique occurrence records, as models produced with small sets of occurrence records tend to underperform and/or produce highly variable predictions, and thus produce misleading predictions of the environmental niche of the taxon under analysis (Pearson *et al.*, 2007; Wisz *et al.*, 2008).

The native range of each crop gene pool, defined as the group of countries where the wild relatives of a crop gene pool have been reported to be native to, was used to determine the background extent of each niche model. Ten thousand random points were created within this area and used as background points to enable each MaxEnt model to characterize the environmental conditions of the area of study. All models were trained using 75% of the occurrence data,

Table 6.1: Full list of general circulation models (GCMs) from the CMIP5 used to project the environmental niches of crop wild relatives. Downloaded from: `www.ccafs-climate.org`

Modelling centre	**Institute ID**	**Model name**	**Country**
Beijing Climate Center, China Meteorological Administration	BCC	BCC-CSM1.1 BCC-CSM1.1(m)	China
College of Global Change and Earth System Science, Beijing Normal University	GCESS	BNU-ESM	China
Canadian Centre for Climate Modelling and Analysis	CCCma	CanESM2	Canada
Community Earth System Model Contributors	NSF-DOE-NCAR	CESM1(CAM5) CESM1(BGC)	USA
Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM)	CSIRO-BOM	ACCESS1.0 ACCESS1.3	Australia
Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	CSIRO-QCCCE	CSIRO-Mk3.6.0	Australia
The First Institute of Oceanography, SOA	FIO	FIO-ESM	China
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-CM3 GFDL-ESM2G GFDL-ESM2M	USA
NASA Goddard Institute for Space Studies	NASA GISS	GISS-E2-H-CC GISS-E2-R GISS-E2-R-CC	USA
Institute for Numerical Mathematics	INM	INM-CM4	Russia
Institut Pierre-Simon Laplace	IPSL	IPSL-CM5A-LR IPSL-CM5A-MR	France
LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University	LASG-CESS	FGOALS-g2	China
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	MIROC	MIROC-ESM MIROC-ESM-CHEM	Japan
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC	MIROC5	Japan
Met Office Hadley Centre and Instituto Nacional de Pesquisas Espaciais	MOHC	HadGEM2-CC HadGEM2-ES	UK
Max Planck Institute for Meteorology	MPI-M	MPI-ESM-LR	Germany
Meteorological Research Institute	MRI	MRI-CGCM3	Japan
National Center for Atmospheric Research	NCAR	CCSM4	USA
Norwegian Climate Centre	NCC	NorESM1-M	Norway
National Institute of Meteorological Research/Korea Meteorological Administration	NIMR/KMA	HadGEM2-AO	Korea

and tested using the remaining 25% of the occurrence data available for each crop wild relative taxa. Default settings in MaxEnt were used for producing the environmental niche models. Once each model was trained and tested, they were projected into the current and future environmental layers. Binary maps for each projection were produced by applying the "Maximum training sensitivity plus specificity" threshold, as it minimizes commission (overprediction) and ommission (underprediction) errors (Liu *et al.*, 2005). The area under the receiver operating (AUC) was used as the parameter to assess the performance of environmental niche models. An AUC value above 0.7 was considered adequate for this study, as models with an AUC value above 0.7 are considered to be able to discriminate the areas where a species is likely to occur.

Binary maps with AUC > 0.7 were used to estimate suitable areas for crop wild relatives under current and future climates.

6.3.5 Impacts

Two contrasting adaptation scenarios were considered for this study: One assuming that plant species are not able to track their climates and thus no potential habitat gains are possible (i.e., zero-dispersal scenario), and a more optimistic scenario where it is assumed that plants are able to track their climates within their native ranges (i.e., optimistic dispersal scenario). The optimistic dispersal scenario does not take into account other parameters that may affect the movement of plant species like the interactions with competitors or the alterations in the land cover, nor it assumes a particular speed of dispersal (e.g., Warren *et al.*, 2013).

For the zero-dispersal scenario, the future environmental niche models were compared to the environmental niche models produced with the baseline information, and the loses of current suitable areas were computed. For the optimistic dispersal scenario, the native ranges of each taxa are used to clip the future and current environmental niches. Information on native ranges was obtained from the GRIN Taxonomy portal (USDA ARS National Genetic Resources Program, n.d.). Once the current and future environmental niches are clipped, they are compared and the projected gains and losses in suitable area are computed.

6.3.6 Taxa richness

I mapped the taxa richness under current and future climatic conditions by overlapping the individual distribution models of taxa with adequate models (AUC > 0.7). This was done with the aim of identifying global patterns of richness of these wild relatives, likely changes under future climatic conditions, and identification of sites where potential gains and loses are likely to take place. Overall changes in the composition of wild relative taxa in the space (i.e., taxa turnover) were calculated by comparing the proportion of projected gains and loses against the current distributions of taxa (Peterson *et al.*, 2015).

6.4 Results

6.4.1 Crops and species selection

A total of 432 unique wild relative taxa, associated to 29 crops were selected (Supplementary Table 6.1). Eight botanical families and 26 genera are represented in the selection of crops and wild relative taxa. 22% of the wild relative taxa selected are classified in the primary gene pool concept, 56% are in the secondary gene pool, and 21% are in the tertiary gene pool. 1% taxa are in the Taxon Group 1B, and less than 1% are in the Taxon Group 2 or have not been classified in either category yet.

6.4.2 Occurrence data

A total of 136,003 occurrence records with geographical coordinates were retrieved, with some taxa having up to 6,407 records with coordinates and 125 taxa with less than ten coordinates. The overall distribution of the wild relative taxa occurs across the terrestrial portion of the world (including islands), with particular regions displaying a rather high density of occurrence records such as Central America, the Andes, the Fertile Crescent, South East Asia, and Australia.

6.4.3 Environmental niche modelling

As to the environmental drivers used for modelling the environmental niches of wild relative taxa, the Variance Inflation Factor (VIF) detected high correlation between ten environmental predictors. Such highly correlated variables were removed in a stepwise approach, which consisted on removing one of the variables with VIF > 10 and then repeating the procedure until no variables with a VIF > 10 were left. The variables selected to be used as predictor variables in MaxEnt were: BIO 2, BIO 3, BIO 8, BIO 9, BIO 13, BIO 14, BIO 15, BIO 18, BIO 19 and ALT (Table 6.2).

Nearly 70% of the wild relative taxa selected (296 taxa) were modeled as they presented

Table 6.2: List of environmental drivers used in the analysis. VIF is the acronym for Variance Inflation Factor. VIF$_R$ is the Variance Inflation Factor obtained after removing highly correlated variables.

Type	ID	Description	Units	VIF	VIF$_R$
Temperature	BIO 1	Annual mean temperature	°C	1287.95	Excluded
Temperature	BIO 2	Mean diurnal range	°C	20.41	2.7
Temperature	BIO 3	Isothermality	%	12.35	5.05
Temperature	BIO 4	Temperature seasonality	%	2434.6	Excluded
Temperature	BIO 5	Maximum temperature of warmest month	°C	Inf	Excluded
Temperature	BIO 6	Minimum temperature of coldest month	°C	Inf	Excluded
Temperature	BIO 7	Temperature annual range	°C	Inf	Excluded
Temperature	BIO 8	Mean temperature of wettest quarter	°C	9.4	4.19
Temperature	BIO 9	Mean temperature of driest quarter	°C	22.56	3.64
Temperature	BIO 10	Mean temperature of warmest quarter	°C	2089.85	Excluded
Temperature	BIO 11	Mean temperature of coldest quarter	°C	6650.66	Excluded
Rainfall	BIO 12	Annual precipitation	mm	72.1	Excluded
Rainfall	BIO 13	Precipitation of wettest month	mm	89.06	5.16
Rainfall	BIO 14	Precipitation of driest month	mm	94.26	3.81
Rainfall	BIO 15	Precipitation seasonality	adimensional	3.31	2.33
Rainfall	BIO 16	Precipitation of wettest quarter	mm	152.77	Excluded
Rainfall	BIO 17	Precipitation of driest quarter	mm	122.11	Excluded
Rainfall	BIO 18	Precipitation of warmest quarter	mm	5.51	4.43
Rainfall	BIO 19	Precipitation of coldest quarter	mm	4.03	3.52
Elevation	ALT	Elevation	m	2.68	1.92

more than ten records with unique geographical coordinates. From this, 290 taxa were considered for the climate change analysis as the AUC values of their models were above 0.7. Six taxa were excluded from the analysis (AUC < 0.7). Models for the wild relatives of chickpea and faba bean did not meet the conditions of the AUC parameter or number of unique georeferenced occurrences required to construct environmental niche models.

6.4.4 Impacts

The wild relatives of finger millet (-34.0%), potato (-32.2%) and cowpea (-33.28%) appear to be the most affected under all dispersal scenarios due to the high average loss of suitable area under the future conditions of climate change when compared to the rest of the crop gene pools analyzed (Table 6.3). Crop gene pools of vetch, pea, pearl millet, carrot and oat present a rather high percentage of loses (22 - 26%). However, large variations within each gene pool suggests that only some wild relatives will be largely affected by climate change. In contrast, the crop gene pools of rice (-7.1%) and pigeonpea (-7.8%) appear to be less susceptible to the projected changes in climates (Figure 6.2, Table 6.3).

The effect of climate change is larger under the no-disperal scenario, as it assumes that wild

Table 6.3: Mean effect of climate change on climatically suitable areas of wild relative taxa

		No-dispersal scenario		Optimistic dispersal scenario	
Crop name	**No. of crop wild relative taxa**	**No. of taxa losing > 50% range**	**Average % change (SD)**	**No. of taxa losing > 50% range**	**Average % change (SD)**
African rice (*Oryza glaberrima*)	20	0	-7.07 (10.93)	0	27 (33.46)
Alfalfa (*Medicago sativa*)	12	0	-8.38 (3.96)	0	5.62 (15.93)
Apple (*Malus domestica*)	30	0	-14.23 (13.35)	0	14.03 (20.09)
Asian rice (*Oryza sativa*)	20	0	-7.07 (10.93)	0	27 (33.46)
Bambara groundnut (*Vigna subterranea*)	2	0	-10.32 (0)	0	-2.59 (0)
Banana (*Musa acuminata*)	6	1	-19.48 (21.09)	0	-15.61 (23.96)
Barley (*Hordeum vulgare*)	4	0	-8.65 (4.22)	0	11.99 (24.33)
Bean (*Phaseolus vulgaris*)	7	0	-10.95 (13.11)	0	6.39 (33.69)
Carrot (*Daucus carota*)	21	1	-24.09 (17.85)	1	1.28 (32.33)
Chickpea (*Cicer arietinum*)	5	-	-	-	-
Cowpea (*Vigna unguiculata*)	14	2	-33.28 (28.89)	2	-19.11 (41.08)
Eggplant (*Solanum melongena*)	52	1	-20.61 (20.21)	1	-5.46 (30.69)
Faba bean (*Vicia faba*)	1	-	-	-	-
Finger millet (*Eleusine coracana*)	6	1	-34.04 (19.58)	1	-27.17 (26.29)
Grasspea (*Lathyrus sativus*)	5	0	-13.39 (11.89)	0	0.26 (22.97)
Lentil (*Lens culinaris*)	5	0	-16.91 (9.49)	0	-9.77 (12.81)
Lima bean (*Phaseolus lunatus*)	4	0	-7.49 (5.79)	0	0.73 (12.2)
Oat (*Avena sativa*)	15	2	-22.75 (17.23)	2	-4.9 (35.3)
Pea (*Pisum sativum*)	5	0	-24.99 (20.07)	0	-16.79 (27.16)
Pearl millet (*Pennisetum glaucum*)	5	0	-24.22 (23.51)	0	-18.14 (21.82)
Pigeonpea (*Cajanus cajan*)	15	1	-7.82 (20.29)	1	37.44 (51.45)
Plantain (*Musa balbisiana*)	6	1	-19.48 (21.09)	0	-15.61 (23.96)
Potato (*Solanum tuberosum*)	73	15	-32.19 (23.96)	13	-22.85 (32.78)
Rye (*Secale cereale*)	4	0	-17.95 (25.26)	0	-9.62 (23.93)
Sorghum (*Sorghum bicolor*)	17	0	-13.23 (16.97)	0	14.1 (36.73)
Sunflower (*Helianthus annuus*)	36	2	-14.21 (15.48)	2	7.09 (28.29)
Sweet potato (*Ipomoea batatas*)	14	0	-12.9 (10.49)	0	8.13 (22.71)
Vetch (*Vicia sativa*)	9	1	-26.23 (21.06)	1	-17.87 (23.49)
Wheat (*Triticum aestivum*)	42	1	-17.81 (15.77)	1	-2.99 (24.08)

relative taxa are not able to respond to climate change by tracking the climate favourable for their growth, development and subsistence. The overall proportion of distribution losses for all the crop gene pools analyzed is of 20.8%, where potato appear to be the most susceptible due to the number of wild relatives losing more than 50% of their distributions (Table 6.4). Moreover, seven of these susceptible potato wild relatives were determined to be poorly represented in genebanks (Castañeda-Álvarez *et al.*, 2015, 2016). One species in particular, *Solanum burkartii*, is under the risk of extinction due to the negative impact of climate change on its suitable area ($\sim$ -90%) (Table 6.4). Other wild relatives projected to be largely affected due to climate change include *Vigna keraudrenii* and *Vigna monantha* (wild relatives of cowpea), *Avena monantha* (wild relative of oat), *Solanum contumazaense* and *Solanum tarnii* (wild relatives of potato) and *Helianthus paradoxus* (wild relative of sunflower) (Table 6.4).

Under the optimistic dispersal scenario the overall changes (gains and losses together) in distributional area compute 26.6%, and average gains reach up to 25% of the current potential

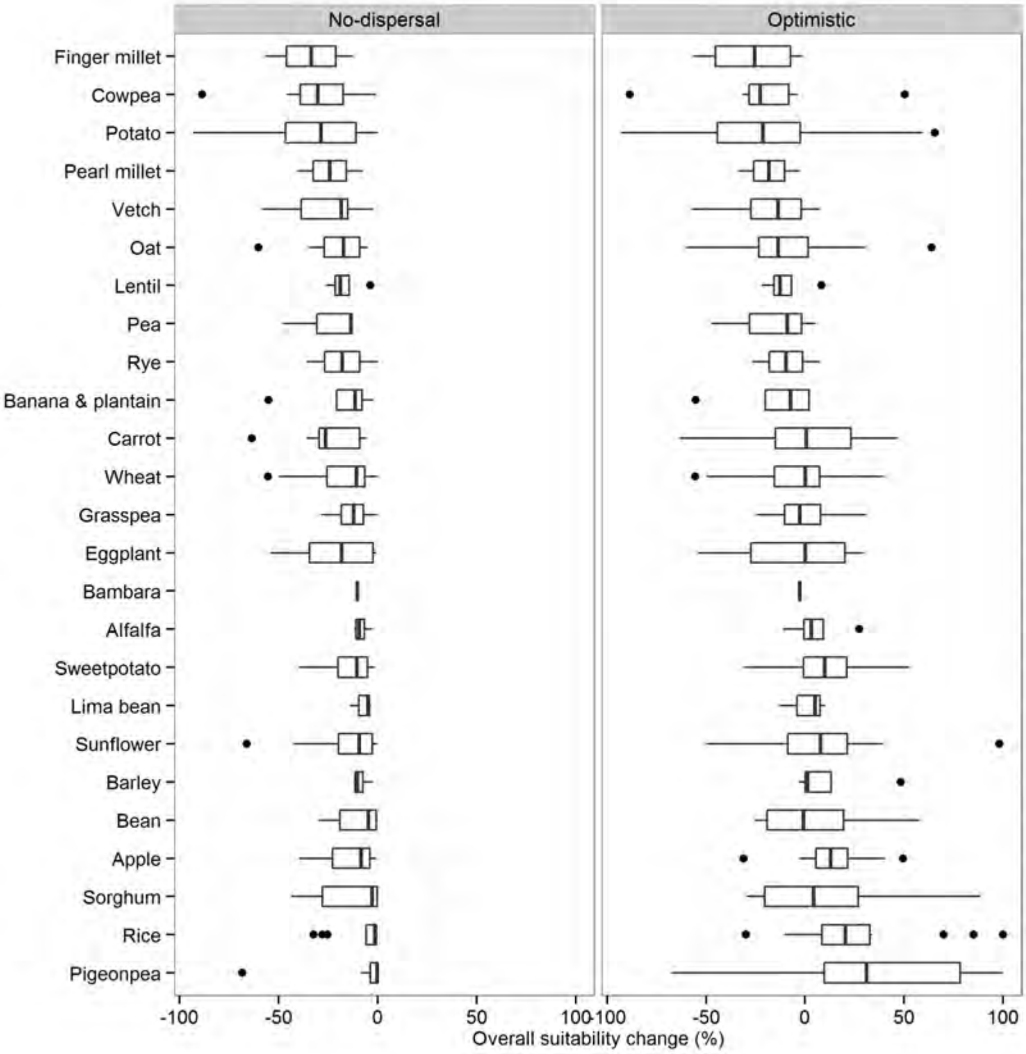


Figure 6.2: Impacts of climate change on the climatic suitable areas of wild relative taxa grouped by crop gene pool. The bold vertical line in each box is the median. Each box displays the first and third quartile. Outliers are plotted as individual points.

areas. The wild relatives of finger millet, cowpea and potato are again the most affected, suggesting that the losses in area are greater than the potential gains that may arise under climate change (Figure 6.2). In contrast, the wild relatives of pigeonpea and rice present average gains of 27% of their current potential distribution. The wild relatives of pigeonpea, rice and apple present higher relative gains among all crops (Figure 6.3).

In terms of the geographic distributions of crop wild relatives, richness hotspots are found in the Mediterranean basin, Near East, Northern Australia, and mountainous regions in Central and South America (Figure 6.4a). Potential gains in suitable area for up to 12 taxa are

Table 6.4: List of most impacted crop wild relatives due to climate change. FPS: Final priority score. Collecting prioritization category and FPS were obtained from (Castañeda-Álvarez *et al.*, 2016).

		% of change in suitable area		
Crop name	**Crop wild relative taxon**	**Non-dispersal scenario**	**Optimistic dispersal scenario**	**Collecting prioritization category (FPS)**
Banana & plantain	*Musa yunnanensis*	-55.22	-55.19	High (8.3)
Carrot	*Daucus carota* subsp. *halophilus*	-63.63	-63.47	High (9.8)
Cowpea	*Vigna keraudrenii*	-88.63	-88.58	High (10)
Cowpea	*Vigna monantha*	-87.41	-87.41	High (10)
Eggplant	*Solanum lichtensteinii*	-54.34	-54.32	High (9.4)
Finger millet	*Eleusine floccifolia*	-56.71	-56.65	High (9)
Oat	*Avena maroccana*	-60.35	-59.93	Low (3.4)
Oat	*Avena murphyi*	-85.58	-85.58	Medium (6.5)
Pigeonpea	*Cajanus acutifolius*	-68.46	-67.63	High (8.5)
Potato	*Solanum burkartii*	-93.02	-92.92	High (10)
Potato	*Solanum contumazaense*	-88.52	-86.17	High (10)
Potato	*Solanum garcia-barrigae*	-70.66	-69.96	High (10)
Potato	*Solanum gracilifrons*	-66.15	-64.77	High (10)
Potato	*Solanum hastiforme*	-55.73	-54.86	High (10)
Potato	*Solanum hjertingii*	-66.62	-66.16	High (7.2)
Potato	*Solanum lesteri*	-63.27	-60.02	Medium (6)
Potato	*Solanum morelliforme*	-51.73	-50.62	Medium (5.9)
Potato	*Solanum olmosense*	-63.73	-49.92	High (10)
Potato	*Solanum paucissectum*	-64.79	-55.76	Low (3)
Potato	*Solanum polyadenium*	-75.43	-75.09	Medium (5.3)
Potato	*Solanum sogarandinum*	-63.92	-61.67	Medium (6.2)
Potato	*Solanum tarnii*	-82.61	-81.68	Medium (6.3)
Potato	*Solanum vernei*	-50.28	-43.26	Low (5)
Potato	*Solanum verrucosum*	-64.94	-62.36	Medium (5.2)
Sunflower	*Helianthus niveus* subsp. *tephrodes*	-66.12	-50.18	High (8.9)
Sunflower	*Helianthus paradoxus*	-94.91	-94.91	Low (3.8)
Vetch	*Vicia sativa* subsp. *amphicarpa*	-58.53	-57.32	Low (5)
Wheat	*Aegilops searsii*	-55.46	-55.41	No urgent collecting required (1.2)

found scattered over the world (Figure 6.4b), while the potential losses in areas for 1-5 taxa are localized in temperate regions and tropical rainforests (e.g., South America and Africa) with significant losses displayed in Western Europe (11-15 taxa) (Figure 6.4c).

High taxa turnover, or changes in the composition of crop wild relative taxa, are projected in the limits of the distributions of crop wild relatives (e.g., temperate regions), Western Europe and the Amazonian basin (Figure 6.4). Coinciding with regions where large losses of suitable area are expected. No large turnover is expected in some hotspots of crop wild relative taxa richness, such as the Andes, Central America, parts of the Near East and Northern Australia.

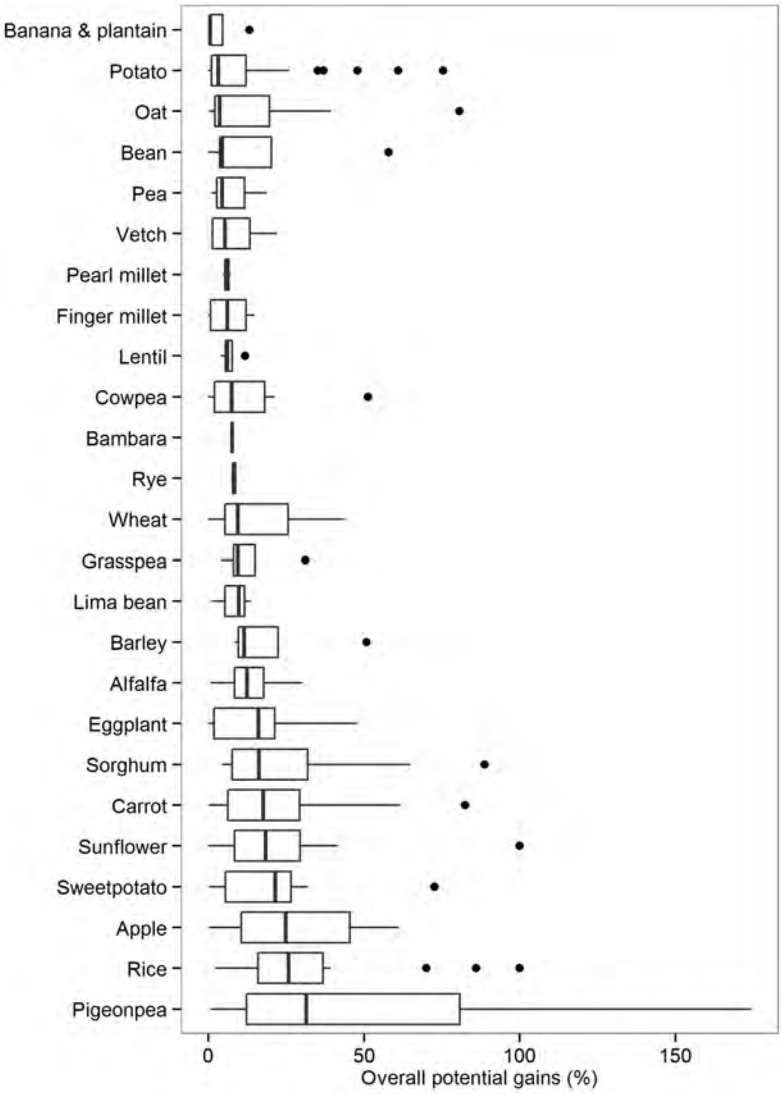


Figure 6.3: Potential wild relatives range gains grouped by crop gene pool under an optimistic dispersal scenario. The bold vertical line in each box is the median. Each box displays the first and third quartile. Outliers are plotted as individual points.

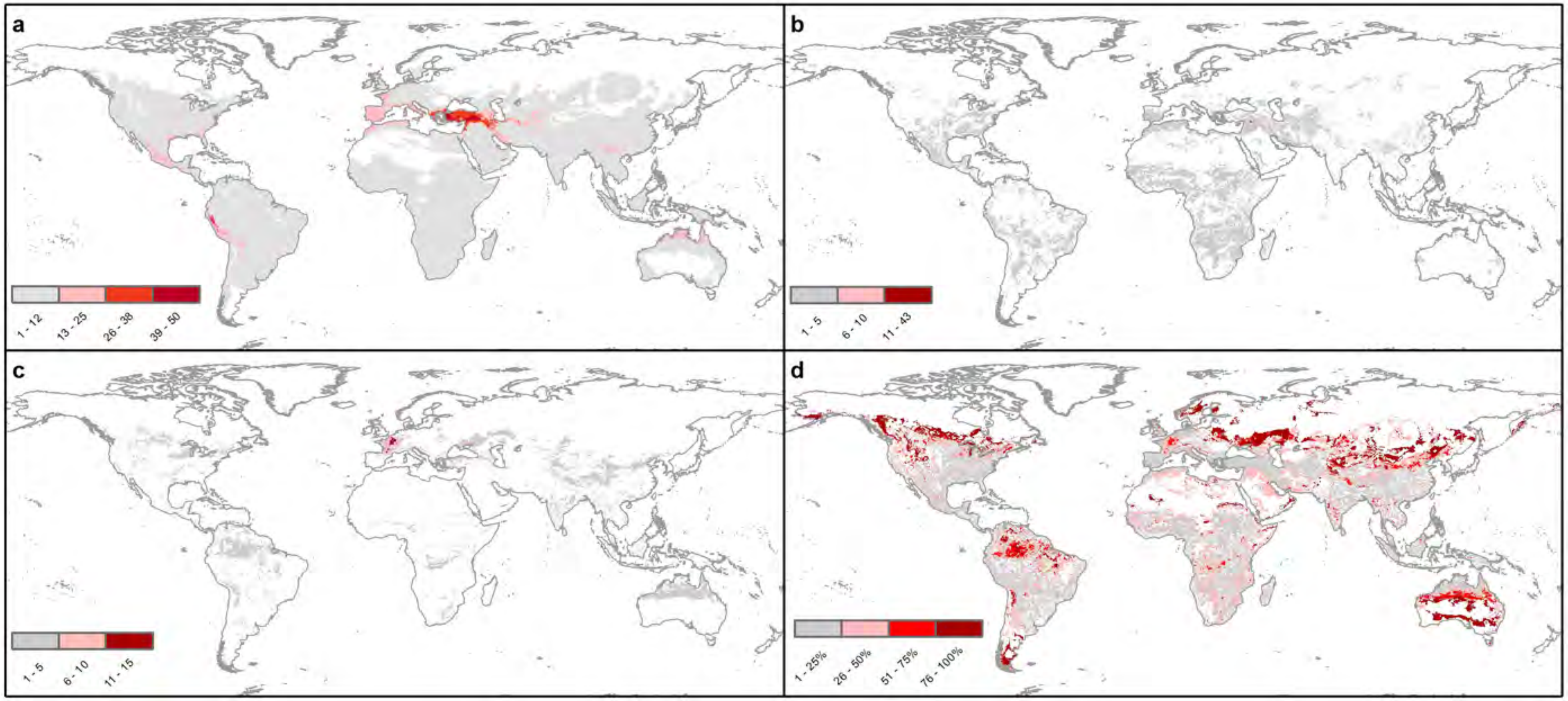


Figure 6.4: Modelled crop wild relative taxa richness patterns and climate change impacts. **a** Current potential taxa richness: gray, 1-12 taxa; pink, 13-25 taxa; red, 26-38 taxa; dark red, 39-50 taxa. **b** Potential gains: gray, 1-5 taxa; pink, 6-10 taxa; dark red, 11-43 taxa. **c** Potential losses: gray, 1-5 taxa; pink, 6-10 taxa; dark red, 11-15 taxa. **d** Taxa turnover: gray, 1-25%; pink, 26-50%; red, 51-75%; dark red, 75-100%.

6.5 Discussion

The goal of this study was to assess the potential impacts of climate change on the distributions of crop wild relative taxa across multiple crops. Analysis of changes in the suitable areas of crop wild relatives under a stringent emissions scenario (RCP 4.5) shows an average of area loses of 20.8%. Nearly 10% of the taxa assessed are projected to lose more than 50% of their current distributions. Geographic regions likely to be most impacted by the changes in suitable areas for crop wild relatives include temperate zones, and the neotropics.

Some crop gene pools (i.e., the group of wild relative taxa of a crop) appear to be more negatively impacted by climate change, as it is the case for finger millet, cowpea and potato. In the same way, some taxa are projected to be more affected than other, reaching the risk of extinction due to large losses in suitable area, as it is the case of *Vigna keraudrenii*, *Vigna monantha*, *Avena murphyi*, *Solanum burkartii*, *Solanum contumazaense*, *Solanum tarnii* and *Helianthus paradoxus).*

In a previous study, Jarvis *et al.* (2008) by using a different emissions scenario and modelling algorithm estimated rather large losses in area suitability when compared with those reported in this study. Jarvis *et al.* (2008) report losses between 38 - 69% of the distribution areas of potato, gains of 8% and losses of 65% for cowpea. The large differences in average losses between both studies are probably influenced by the number and taxonomic identity of crop wild relative taxa included for each crop gene pool in each study. However, at the taxa level, both studies agree in identifying *Vigna keraudrenii*, *Solanum paucissectum* and *Solanum tarnii* as taxa highly impacted by climate change.

Investments aiming to conserve the wild relatives of crops need to consider the likely impacts of climate change for durable and long-term conservation. The results presented in this study vary between and within crop gene pools, and between dispersal scenarios. Different conservation strategies are needed according to the impact level of climate change on the geographical distributions of crop wild relatives.

From a taxa level approach, wild relatives with large negative impacts on their distributions under both dispersal scenarios may require to be better represented and conserved in *ex situ* genebanks due to the significant reduction of their environmentally suitable areas. In the case of wild relative taxa presenting minor loses in their future distribution, they could be considered pertinent candidates for *in situ* conservation due to the relative stability of their environmental niches. However, complementary *ex situ* conservation is desirable as it can help easing the access of these resources for plant breeding and research. Monitoring the genetic diversity and responses to climate change of wild relative taxa with minor loses and/or potential gains will also provide critical information on the responses and changes in the genetic composition of these plants' populations. Additionally, monitoring new competitors and the effect of habitat fragmentation and land use changes can also help understanding if the taxon can effectively occupy the projected territorial gains.

In terms of setting priorities for conservation, taxa and crop gene pools displaying the largest losses in the future should be prioritized for conservation work, due to their high vulnerability to climate change. Additional variables can be also considered to determine the prioritization urgency of taxa and crop gene pools, as the economic value of the associated crops, the nutritional supply of the associated crops to human diets, the current conservation status (e.g., IUCN Red List), the current extent of representativeness of the genetic diversity of such taxa and crop gene pools in existing *ex situ* holdings and *in situ* sites. Taxa such as *Musa yunnanensis* and *Daucus carota* subsp. *halophilus* (see Table 6.4 for a complete list of taxa) have been already targeted as being poorly represented in genebanks (Castañeda-Álvarez *et al.*, 2016), increasing the urgency of improving their conservation and representativeness in *ex situ* holdings.

Prioritizing climatically vulnerable crop wild relatives should be also guided by the easiness of crossing such wild taxa with their associated crops (i.e., taxa in the primary and secondary gene pools or equivalent categories), as the production of fertile hybrids is expected to be less cumbersome. Such prioritization should also take into account less related taxa (i.e., third gene pool and equivalents) that have been already used in plant breeding as sources of genetic diversity and specific traits.

From a geographical perspective, areas harboring high taxa richness and presenting minor to no changes in their future taxa richness should be considered as candidate sites for establishing *in situ* conservation initiatives thanks to the climatic stability they display (e.g., mountainous regions in Central and South America, the Iberian Peninsula, and part of the Near East). Whereas, areas with large losses of richness in the future might be considered less adequate for establishing *in situ* reserves for long-term conservation, and therefore efforts to secure this genetic diversity in genebanks are warranted. Given the lack of information on the physiological and adaptive response of wild relative taxa to climate change, areas detected to have major losses should be sampled for subsequent conservation in genebanks (e.g., temperate zones and Amazonian basin).

I acknowledge the limitations of the approach used in this study, as it only takes into account the environmental suitability of the wild relative taxa analyzed under a stringent representative concentration pathway, and two highly contrasting species adaptation scenarios. Further studies, using a similar approach to that presented here, but considering more than one representative concentration pathway and multiple periods of time can eventually provide estimates of the consequences of implementing (or not) actions towards the reduction of emissions at a given point in time (e.g., Warren *et al.*, 2013). Additionally, using multiple environmental niche modelling algorithms for producing the environmental niches of wild relative taxa in the future can serve as a way to estimate and inform uncertainty (Araújo and New, 2007).

Further elements to investigate the impact of climate change on the wild relatives of crops include field experiments to identify the likely competitors that crop wild relatives might encounter in new geographic ranges (if dispersal is possible) (Alexander *et al.*, 2015), assessments of the effect of climate change on the interactions of crop wild relatives with beneficial species such as pollinators and animals involved in seed dispersal, and evaluations to determine the type of responses that crop wild relatives may display under future warming conditions. Also, given the massive and rapid transformation of natural habitats to other land uses (e.g., agriculture and urbanization), it is necessary to understand the effects that these land uses may have in the ability of wild relatives to track their climates and occupy new suitable areas for their growth and reproduction.

CHAPTER 7

COMPLEMENTARY DIMENSIONS FOR REFINING GLOBAL CONSERVATION PRIORITIES FOR CROP WILD RELATIVES

7.1 Summary

Agriculture needs to continue producing food for an expanding human population, using less natural resources and reducing its negative impacts to the environment. Moreover, climate change is expected to severely affect crop yields by mid-century (Challinor *et al.*, 2014). The sustainable intensification of agriculture has been proposed as a strategy to increase yields, respond to climate change, use natural resources more efficiently and diminish the footprint of agriculture to the environment (Tilman *et al.*, 2011; Foley *et al.*, 2011). Attaining a more sustainable and intensified agriculture requires new plant varieties able to respond to biotic and abiotic stresses, use resources more efficiently, and able to sustain or increase productivity. Plant breeders can obtain plant varieties containing novel traits thanks to the increasing availability of tools for genomic-based breeding, but this largely depends on the availability of novel sources of genetic diversity such as crop wild relatives (Tanksley and McCouch, 1997; Abberton *et al.*, 2015).

Crop wild relatives, a class of plant genetic resources, are more genetically diverse than cultivars and landraces, and serve as valuable sources of novel traits for agriculture (Tanksley and McCouch, 1997; Mariac *et al.*, 2006). Their conservation, availability and accessibility are critical for producing new plant varieties able to respond to the challenges that agriculture faces (Guarino and Lobell, 2011; McCouch *et al.*, 2013). Despite their importance, crop wild relatives are largely underrepresented in *ex situ* holdings and only few are thoroughly conserved *in situ* reserves (Chapter 5; Maxted and Kell, 2009). Moreover, the resources available for securing crop wild relatives are scarce, and therefore means to set conservation priorities are needed.

Here we present an approach for identifying *in situ* and *ex situ* conservation priorities at the global and more regional scales, based on the value of crop gene pools (i.e., group of wild relative taxa of a crop). For this we estimated the value of crop gene pools using information on global importance of crops to diets and agricultural production systems, and likelihood of use of crop wild relatives in plant breeding. A further refinement of the conservation priorities identified in Chapter 5 is performed by including the crop gene pool values here estimated. Cassava was identified as the crop gene pool of highest priority for conservation because of its importance for global diets, agricultural production systems, high likelihood of being used in plant breeding and low representativeness in genebanks.

7.2 Introduction

Human population is projected to grow to 9.6 billion people by 2050 (United Nations, 2015), almost an additional quarter of the current living population. Despite the significant reduction in the number of undernourished people during the last decade, one in nine people in the world are still undernourished today, and regions such as Southern Asia and sub-Saharan Africa still present high rates of undernourished inhabitants (FAO *et al.*, 2015). At the same time climate change is expected to reduce crop production by 2% per decade (Porter *et al.*, 2014), unless mitigating actions are taken. Producing nutritious food for an increasing human population, while mitigating the likely impact of climate change and reducing agriculture's negative impacts to the environment are required if a more sustainable agriculture is to be achieved. Multiple strategies have been proposed to face the dilemma of producing more food using less natural resources, including reducing the consumption of animal-derived products in human diets, reducing food waste, and increasing yields in established croplands (Foley *et al.*, 2011; Tilman and Clark, 2014). The latter strategy will require, among others, the development and availability of crop varieties that also help abating the projected negative impacts of climate change on agriculture (Lobell *et al.*, 2008; Porter *et al.*, 2014; Asseng *et al.*, 2015).

Crop wild relatives, the wild cousins of cultivated species, are important for agriculture due to their wide genetic diversity and novel traits that can be used to produce new crop varieties (Hajjar and Hodgkin, 2007). Various crop wild relatives' genes have been successfully introgressed to cultivated species, resulting in novel varieties with improved yields (Xiao *et al.*, 1996; Gororo *et al.*, 2002), better nutritional quality (Paulsen, 2000; Traka *et al.*, 2013; Chae *et al.*, 2015), increased tolerance to pest and diseases (Brar and Khush, 1997; Moss *et al.*, 1997; Yadav *et al.*, 2004), and greater resilience to adverse environmental conditions (Guarino and Lobell, 2011; McCouch *et al.*, 2013). The use of crop wild relatives in plant breeding is expected to continue to grow as more of the traits they harbor are identified (Sarikamis *et al.*, 2006; Glover *et al.*, 2010), the mechanisms of inter-species hybridization are improved (Lin *et al.*, 2015), and pre-bred material is made readily available to plant breeders and researchers (Dempewolf *et al.*, 2013).

To continue and expand their use, crop wild relatives need to be conserved and available to plant breeders and users in general. Genebanks are facilities conserving and facilitating access to plant genetic resources, including the wild relatives of crops. Currently, crop wild relatives are poorly represented in genebanks. Only 10% of all the plant germplasm accessions visible to the global research and plant breeding community are crop wild relatives, and seven in ten crop wild relatives' taxa associated to the world's most important crops have been accounted to be un- and/or under-represented in genebanks (Castañeda-Álvarez *et al.*, 2016). Furthermore, *in situ* conservation of crop wild relatives is very limited, protecting only a few taxa in some regions (Maxted and Kell, 2009). Such reserves are also in need of implementing quality standards to ensure the successful and long-term conservation of wild relative taxa (Iriondo *et al.*, 2012). Targeted conservation can help using financial resources efficiently to improve the representativeness of crop wild relatives in genebanks, and prioritize areas for *in situ* conservation.

Improving the conservation of crop wild relatives also contributes to global policy targets, specifically: Goal 2 of the United Nation's Sustainable Development Goals, and Target 13 of the "Aichi Biodiversity Targets" (Secretariat of the CBD, 2010).

Various approaches have been developed and applied to determine priorities for conservation, and subsequently translated into targeted conservation actions. These approaches prioritize crop wild relatives at the taxa level (e.g., species risk of extinction, genetic erosion) and crop-level metrics (e.g., economic value of associated crop, contributions of associated crop in human diets), producing different listings which would be useful in different contexts and scales. Flor *et al.* (2006) prioritized European crop wild relative taxa requiring conservation by assessing their risk of extinction, their *in situ* and *ex situ* conservation status, the degree of relatedness of wild relatives to their associated crops, the genetic erosion and/or pollution of wild relatives, the actual and potential economic importance of crop wild relatives at the national and regional level, and the cultural uses of wild relatives. Ford-Lloyd *et al.* (2008) proposed a pragmatic approach, by assessing one of these parameters: the extent of distribution of wild relative taxa in their natural habitats, or the socio-economic value of crop wild relatives obtained by consulting experts in the field. Brehm *et al.* (2010) defined the conservation priorities of crop wild relatives of Portugal, by investigating whether the wild relative taxa were native or not to the country, the economic and ethnobotanic value of wild relative taxa, the extent of global and local distribution of the wild relative taxa, the *in situ* and *ex situ* conservation status, existence of national legislation protecting the taxa, and IUCN Red Listing category.

Recently, the global hotspots of wild relative taxa richness and the major *ex situ* and *in situ* conservation gaps have been identified by examining the extent of diversity represented in genebanks (Chapter 5) and protected areas (H. Vincent, personal communication). However, basing future *ex situ* collecting or *in situ* site designation on the extent of diversity represented in genebanks and/or *in situ* sites alone assumes all different crop wild relatives have an equal value. Moreover, given the limited availability of resources for conservation, initial conservation action should focus on crop wild relative taxa of highest potential utilization value. Value in this sense may be defined in terms of the relative commercial and macronutrient value of associated crops and the relative closeness of crop wild relatives to the crop, and therefore its likelihood of being used in plant breeding. The results presented here are expected to help prioritize future conservation efforts for 81 globally important crops, not only based on the gaps in *ex situ* collections, but also taking into account the value of crop wild relative taxa.

7.3 Methodology

7.3.1 Selection of associated crops and their wild relative taxa

Crops were selected according to their global importance for food security and smallholder income generation. All the food crops covered under the Multilateral System of the International Treaty on Plant Genetic Resources for Food and Agriculture, an international agreement that facilitates access to and use of some plant genetic resources (FAO, 2009), were included in this study.

The group of crop wild relative taxa associated to a crop is called "crop gene pool". The gene pool concept (Harlan and de Wet, 1971) was used to identify the degree of relatedness between wild relative taxa and their corresponding associated crop. The gene pool concept provides information on the easiness of crossing wild relative taxa with their associated crop, and the resulting chances of producing viable offspring. Proxy gene pool concepts were used when no gene pool concept based on crossing data was available, namely, the Taxon Group concept (Maxted *et al.*, 2006), the genetic relative status (Wiersema *et al.*, 2012), and the preliminary gene pool concept (Vincent *et al.*, 2013). Crop wild relative taxa in the primary and secondary gene pool categories (and equivalent categories in proxy concepts) were selected for this study. Less related taxa were included when confirmed and potential uses in plant breeding were reported in the literature (e.g., USDA ARS National Genetic Resources Program, n.d.). All wild relative and associated crop relatedness degree information were obtained by querying The Harlan and De Wet Crop Wild Relative Inventory (`http://www.cwrdiversity.org/checklist/`; Vincent *et al.*, 2013), and GRIN Taxonomy for Plants (USDA ARS National Genetic Resources

Program, n.d.).

7.3.2 Gathering and preparation of occurrence data

Occurrence data of crop wild relative taxa were acquired by querying digital repositories; digitizing herbarium vouchers, peer-reviewed and gray literature, and experts' archives (Chapter 2). All records without geographical coordinates but with detailed descriptions of the locality where the taxa were collected and/or reported were georeferenced using GEOLocate (Rios and Bart, 2010) and the Google Maps Geocoding API. Only the newly calculated coordinates with a precision distance below 5 km were temporarily assigned to the corresponding occurrence record. Both newly calculated and existing geographic coordinates were then validated by overlapping the occurrence records on a global administrative map to detect inconsistencies between the administrative unit (e.g., country) described in the occurrence record and the country where the coordinates were mapped (Hijmans *et al.*, 1999). The nomenclatural identity (scientific name) of each record was standardized following established plant species checklists and verified to resolve misspellings or errors. The name contained in each record was queried in GRIN Taxonomy - a widely-recognized checklist of crop wild relatives scientific names (USDA ARS National Genetic Resources Program, n.d.), the Taxonomic Name Resolution Service - an online tool querying multiple checklists such as Tropicos, USDA Plants, Global Compositae Checklist and NCBI Taxonomy (Boyle *et al.*, 2013), and TaxonStand - an R package to standardize scientific names following The Plant List (Cayuela *et al.*, 2012). Further queries were performed to JSTOR plants (`https://plants.jstor.org`), the International Plant Names Index (IPNI; `http://www.ipni.org`) and the Integrated Taxonomic Information System (ITIS; `http://www.itis.gov`) when the name was not found in GRIN, TNRS, or TaxonStand.

7.3.3 Modelling the distributions of wild relative taxa

MaxEnt was selected as the species distribution modelling algorithm for its ability to produce models using presence-only data as inputs. MaxEnt is widely recognized in the ecology community for its overall performance producing robust spatial distribution models (Elith *et al.*, 2006).

The background extent of each model was determined by overlapping the occurrence records of each taxon on a global map divided in six macroregions (North America, South America, Europe, Asia, Africa and Oceania). The limits of the macroregions containing occurrence records were used as the limits of the background extent. Ten thousand random locations were drawn within the delimited background extent, and used as inputs for training the distribution models together with the occurrence records with valid geographic coordinates, and a set of nineteen bioclimatic variables (Nix, 1986) from the Worldclim database (`http://www.worldclim.org/bioclim`; Hijmans *et al.*, 2005) (Table 7.1).

Table 7.1: List of bioclimatic variables (Nix, 1986) used as environmental drivers in MaxEnt.

Name	Type	Description	Units
BIO 1	Temperature	Annual mean temperature	°C
BIO 2	Temperature	Mean diurnal range	°C
BIO 3	Temperature	Isothermality	%
BIO 4	Temperature	Temperature seasonality	Standard deviation x 100
BIO 5	Temperature	Maximum temperature of warmest month	°C
BIO 6	Temperature	Minimum temperature of coldest month	°C
BIO 7	Temperature	Temperature annual range	°C
BIO 8	Temperature	Mean temperature of wettest quarter	°C
BIO 9	Temperature	Mean temperature of driest quarter	°C
BIO 10	Temperature	Mean temperature of warmest quarter	°C
BIO 11	Temperature	Mean temperature of coldest quarter	°C
BIO 12	Precipitation	Annual precipitation	mm
BIO 13	Precipitation	Precipitation of wettest month	mm
BIO 14	Precipitation	Precipitation of driest month	mm
BIO 15	Precipitation	Precipitation seasonality	Coefficient of variation
BIO 16	Precipitation	Precipitation of wettest quarter	mm
BIO 17	Precipitation	Precipitation of driest quarter	mm
BIO 18	Precipitation	Precipitation of warmest quarter	mm
BIO 19	Precipitation	Precipitation of coldest quarter	mm

Models were trained using the cross-validation option (k=5), as this setting maximizes the use of the occurrence records, especially for taxa with small sets of occurrence records, by splitting the dataset in k number of folds to train and test the model (Merow *et al.*, 2013). Once trained, all models were projected onto bioclimatic layers with a spatial resolution of 2.5 arc-minutes ($\sim$ 5 km x 5 km at the Equator) to produce potential distribution maps. Models were assessed to determine their adequacy of being used to produce crop wild relative taxa richness maps. An adequate model met the following performance criteria: (1) the five-fold average of the test sample Area Under the Receiver Operating Characteristic curve (ATAUC) was greater than 0.7, (2) the standard deviation of the ATAUC for the five folds was lower than 0.15, and (3) the proportion of the potential distribution area where the standard deviation was greater than 0.15 was less than 10%. The shortest distance to the upper left corner of the Receiver Operating Characteristic curve was used as threshold to produce binomial (presence-absence) distribution maps (Liu *et al.*, 2005). Each potential distribution map was further restricted by clipping it to the native distributions described in GRIN Taxonomy. A convex hull (i.e., a polygon produced by joining the outermost occurrence records of a taxon) was produced and used to restrict a potential distribution map when no information on native distributions was available.

7.3.4 Estimating the value of associated crops

The value of the associated crops of wild relative taxa was estimated using three dimensions, namely, macronutrient contribution to the human diet, importance to agricultural systems, and breeding value. Each dimension was then quantified by a range of indicators. The contribution of macronutrients in the human diet was quantified through four indicators: calories (kcal/capita/day), protein (g/capita/day), fat (g/capita/day), and food weight (g/capita/day). The importance to agricultural production systems was quantified using three indicators: production quantity (tonnes), harvested area (ha), and production value (current million USD). For these two dimensions, global aggregated data from FAO statistical averaged over the period 2009-2011 (Food and Agriculture Organization of the United Nations (FAO)., 2013) used in (Khoury *et al.*, 2015c) were used. The third dimension, likelihood of use of wild relatives in plant breeding was produced by assigning a value of three to the wild relative taxa classified in gene pool 1 (GP-1), two to the wild relatives classified in GP-2, and one to the wild relatives in GP-3. Crop wild relative taxa reported as being useful as rootstock and/or graftstock were assigned a value of two, while crop wild relatives with no gene pool classification but reported potential or confirmed use in plant breeding were assigned a value of one. Then all values were added up per crop gene pool. Each dimension was then normalized by dividing each crop's value by the maximum existing value across all crops per dimension. The four macronutrient and three agricultural production indicators were averaged separately. All metrics are in a scale from zero to ten, where zero represents low importance and ten is high importance at the global scale. Finally, the overall value was computed by averaging the three different dimensions.

Associated crops were grouped in four importance categories for each dimension (i.e., macronutrient contributions to human diets, contributions to agricultural production systems, likelihood of use of wild relatives in plant breeding, and overall value) as follows:

- Category "very important": groups dimension values within the third and fourth quartile, including upper outliers (76 - 100%).
- Category "important": includes dimension values between the second and third quartile (51 - 75%).
- Category "moderately important": contains dimension values between the first and second quartile (26 - 50%).
- Category "slightly important": includes dimension values in the first quartile and lower outliers (0 - 25%).

7.3.5 Richness maps per importance categories

Wild relative taxa richness maps were prepared for each crop gene pool by adding the binomial potential distribution models that met the adequacy and performance criteria used in this study. Four global richness maps were prepared for each value dimension by adding the wild relative taxa richness maps of crop gene pools according to the importance categories used to rank crops (i.e., very important, important, moderately important, slightly important). These richness maps were produced to identify conservation hotspots of crop gene pools likely to be used in plant breeding programs, and wild relative taxa of crops contributing to global agricultural production systems and human diets.

7.3.6 Relationships between prioritization scores

The relationship between the crop value dimensions and the final priority score for collecting and conserving crop wild relatives (Final Priority Score in Chapter 5) was examined with a pairwise correlation test, by computing the Spearman's correlation coefficient. A pair of variables were considered complementary when a weak correlation coefficient was estimated ($\rho < 0.5$).

7.4 Results

7.4.1 Crops' aggregation and geographical patterns of crop wild relatives

Cassava, potato, rice and wheat were the only crops classified in the "very important" category for three dimensions, reflecting the global importance of these crops in terms of their contributions to agricultural production systems, and macronutrients to diets (Table 7.2). Likewise, the gene pools of these crops are largely composed by wild relative taxa that are more likely to be used in a plant breeding program thanks to their easiness to cross with the cultivated species, and likelihood of producing fertile offspring (i.e., primary and secondary gene pool or equivalent proxy categories). In terms of the global distribution of wild relative taxa, Central and South America, the Middle East and Southeast Asia emerge as spots of high richness of wild relatives of crop gene pools highly likely to be used in plant breeding, and globally important for food and agriculture (Figure 7.1a).

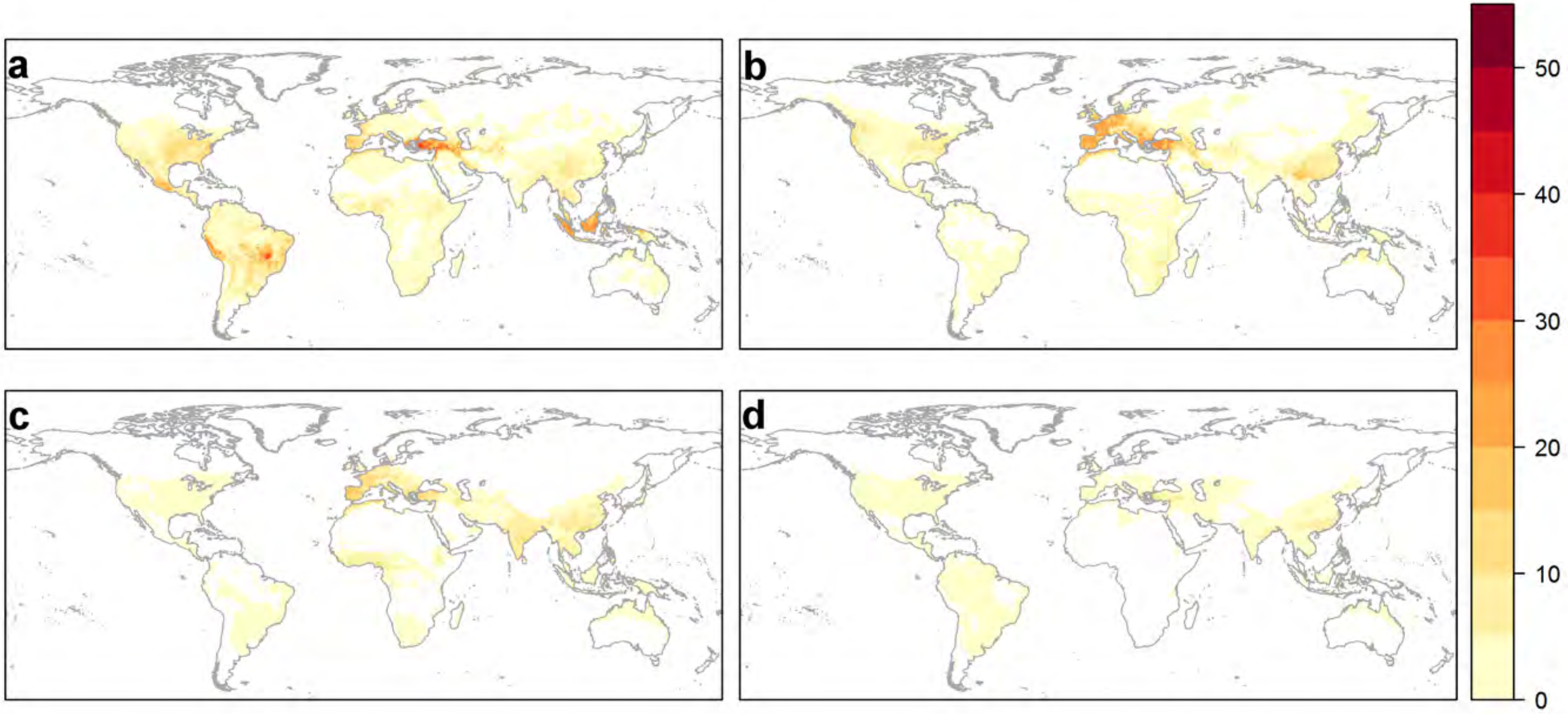


Figure 7.1: Richness maps of crop wild relatives grouped by overall value of associated crops. **a.** Richness map of the wild relatives of crops in the “very important” category; **b.** Richness map of the wild relatives of crops classified in the “important” category; **c.** Richness map of the wild relatives of crops in the “moderately important” category; **d.** Richness map of the wild relatives of crops classified in the “slightly important” category. Darker red colours indicate high overlap of crop wild relative taxa.

Table 7.2: List of crops with importance scores and categories for all crop value dimensions .

Crop type	Crop name (scientific name)	No. of taxa	Crop value dimensions			
			Contributions to human diets (score)	**Importance to agricultural systems (score)**	**Breeding value (score)**	**Overall value (score)**
Pulses	Adzuki bean (*Vigna angularis* (Willd.) Ohwi & H. Ohashi)	12	moderately important (0.09)	slightly important (0.09)	important (0.99)	moderately important (0.39)
Cereals	African rice (*Oryza glaberrima* Steud.)	20	very important (6.62)	very important (7.12)	very important (1.68)	very important (5.14)
Forages	Alfalfa (*Medicago sativa* L.)	17	-	very important (1.05)	very important (1.31)	very important (1.18)
Fruits	Almond (*Prunus dulcis* (Mill.) D. A. Webb)	30	slightly important (0.07)	slightly important (0.12)	very important (2.77)	very important (0.99)
Fruits	Apple (*Malus domestica* Borkh.)	30	important (0.44)	important (0.91)	very important (2.22)	very important (1.19)
Fruits	Apricot (*Prunus armeniaca* L.)	15	slightly important (0.05)	slightly important (0.06)	very important (1.2)	moderately important (0.44)
Cereals	Asian rice (*Oryza sativa* L.)	20	very important (6.62)	very important (7.12)	very important (1.68)	very important (5.14)
Vegetables	Asparagus (*Asparagus officinalis* L.)	16	important (0.34)	slightly important (0.13)	important (1.06)	moderately important (0.51)
Pulses	Bambara (*Vigna subterranea* (L.) Verdc.)	2	moderately important (0.09)	slightly important (0.01)	slightly important (0.19)	slightly important (0.1)
Fruits	Banana & plantain (*Musa acuminata* Colla)	6	very important (0.8)	very important (0.95)	moderately important (0.51)	important (0.75)
Fruits	Banana & plantain (*Musa balbisiana* Colla)	6	very important (0.8)	very important (0.95)	moderately important (0.51)	important (0.75)
Cereals	Barley (*Hordeum vulgare* L.)	4	very important (1.3)	very important (1.36)	slightly important (0.22)	very important (0.96)
Pulses	Bean (*Phaseolus vulgaris* L.)	6	important (0.46)	very important (0.97)	moderately important (0.4)	important (0.61)
Oilcrops	Black mustard (*Brassica nigra* (L.) K.Koch)	6	very important (0.61)	very important (0.99)	moderately important (0.48)	important (0.69)
Fruits	Breadfruit (*Artocarpus altilis* (Parkinson ex F.A.Zorn) Fosberg)	31	-	-	very important (1.2)	very important (1.2)
Cereals	Broom millet (*Setaria italica* (L.) P. Beauv.)	4	important (0.45)	important (0.71)	slightly important (0.37)	moderately important (0.51)
Vegetables	Cabbage (*Brassica oleracea* L.)	23	important (0.34)	important (0.56)	very important (1.68)	very important (0.86)
Fruits	Cacao (*Theobroma cacao* L.)	1	moderately important (0.17)	moderately important (0.26)	slightly important (0.11)	slightly important (0.18)
Vegetables	Carrot (*Daucus carota* L.)	21	important (0.34)	moderately important (0.24)	very important (2.15)	very important (0.91)
Starchy roots	Cassava (*Manihot esculenta* Crantz)	135	very important (0.77)	very important (1.33)	very important (10)	very important (4.04)
Fruits	Cherry (*Prunus avium* (L.) L.)	21	slightly important (0.05)	slightly important (0.09)	very important (1.46)	important (0.53)
Pulses	Chickpea (*Cicer arietinum* L.)	5	moderately important (0.09)	moderately important (0.27)	slightly important (0.3)	slightly important (0.22)
Spices	Chili pepper (*Capsicum annuum* L.)	7	moderately important (0.11)	important (0.47)	moderately important (0.59)	moderately important (0.39)
Oilcrops	Cottonseed (*Gossypium hirsutum* L.)	24	important (0.44)	important (0.81)	very important (1.93)	very important (1.06)
Pulses	Cowpea (*Vigna unguiculata* (L.) Walp.)	14	moderately important (0.09)	moderately important (0.22)	very important (1.42)	important (0.58)
Vegetables	Cucumber (*Cucumis sativus* L.)	2	important (0.34)	important (0.49)	slightly important (0.15)	moderately important (0.33)
Vegetables	Eggplant (*Solanum melongena* L.)	52	important (0.34)	moderately important (0.31)	very important (2.73)	very important (1.13)
Pulses	Faba bean (*Vicia faba* L.)	1	moderately important (0.09)	slightly important (0.08)	slightly important (0.11)	slightly important (0.09)
Cereals	Finger millet (*Eleusine coracana* (L.) Gaertn.)	6	important (0.45)	important (0.71)	moderately important (0.55)	important (0.57)
Cereals	Foxtail millet (*Panicum miliaceum* L.)	6	important (0.45)	important (0.71)	moderately important (0.51)	important (0.56)
Spices	Garlic (*Allium sativum* L.)	1	important (0.34)	moderately important (0.28)	slightly important (0.04)	slightly important (0.22)
Fruits	Grape (*Vitis vinifera* L.)	20	important (0.35)	very important (1.01)	important (0.99)	important (0.78)
Fruits	Grapefruit (*Citrus paradisi* Macfad.)	8	slightly important (0.05)	slightly important (0.06)	slightly important (0.3)	slightly important (0.14)
Pulses	Grasspea (*Lathyrus sativus* L.)	5	moderately important (0.09)	slightly important (0.09)	slightly important (0.37)	slightly important (0.18)
Oilcrops	Groundnut (*Arachis hypogaea* L.)	16	very important (1.36)	important (0.77)	important (1.1)	very important (1.08)

Table 7.2 – continued from previous page

			Crop value dimensions			
Crop type	**Crop name (scientific name)**	**No. of taxa**	**Contributions to human diets (score)**	**Importance to agricultural systems (score)**	**Breeding value (score)**	**Overall value (score)**
Starchy roots	Lagos yam (*Dioscorea cayennensis* Lam.)	6	moderately important (0.25)	important (0.6)	important (0.66)	moderately important (0.5)
Vegetables	Leek (*Allium ampeloprasum* L.)	8	important (0.34)	slightly important (0.02)	important (0.88)	moderately important (0.41)
Fruits	Lemon (*Citrus limon* (L.) Osbeck)	7	slightly important (0.08)	moderately important (0.18)	moderately important (0.4)	slightly important (0.22)
Pulses	Lentil (*Lens culinaris* Medik.)	5	moderately important (0.09)	slightly important (0.11)	moderately important (0.4)	slightly important (0.2)
Vegetables	Lettuce (*Lactuca sativa* L.)	15	important (0.34)	moderately important (0.25)	important (1.1)	important (0.56)
Pulses	Lima bean (*Phaseolus lunatus* L.)	5	important (0.46)	very important (0.97)	moderately important (0.4)	important (0.61)
Cereals	Maize (*Zea mays* L.)	10	very important (2.45)	very important (7.72)	moderately important (0.59)	very important (3.58)
Fruits	Mango (*Mangifera indica* L.)	46	slightly important (0.05)	important (0.5)	very important (2.91)	very important (1.15)
Fruits	Melon (*Cucumis melo* L.)	2	slightly important (0.05)	moderately important (0.23)	slightly important (0.15)	slightly important (0.14)
Pulses	Mung bean (*Vigna radiata* (L.) R. Wilczek)	12	moderately important (0.09)	slightly important (0.09)	important (0.73)	slightly important (0.3)
Oilcrops	Mustard (*Brassica juncea* (L.) Czern.)	4	very important (0.61)	very important (0.99)	slightly important (0.19)	important (0.6)
Cereals	Oat (*Avena sativa* L.)	14	slightly important (0.08)	moderately important (0.24)	important (1.06)	moderately important (0.46)
Vegetables	Onion (*Allium cepa* L.)	3	very important (0.55)	important (0.61)	slightly important (0.26)	moderately important (0.47)
Fruits	Orange (*Citrus sinensis* (L.) Osbeck)	12	moderately important (0.28)	important (0.48)	moderately important (0.51)	moderately important (0.42)
Fruits	Papaya (*Carica papaya* L.)	8	slightly important (0.05)	slightly important (0.08)	moderately important (0.51)	slightly important (0.21)
Pulses	Pea (*Pisum sativum* L.)	5	moderately important (0.14)	moderately important (0.33)	moderately important (0.48)	slightly important (0.32)
Fruits	Peach (*Prunus persica* (L.) Batsch)	26	slightly important (0.05)	moderately important (0.26)	very important (2)	important (0.77)
Fruits	Pear (*Pyrus communis* L.)	27	slightly important (0.05)	moderately important (0.3)	very important (2.08)	important (0.81)
Cereals	Pearl millet (*Pennisetum glaucum* (L.) R. Br.)	5	important (0.45)	important (0.71)	moderately important (0.4)	important (0.52)
Pulses	Pigeonpea (*Cajanus cajan* (L.) Millsp.)	15	moderately important (0.09)	slightly important (0.1)	important (0.99)	moderately important (0.39)
Fruits	Pineapple (*Ananas comosus* (L.) Merr.)	5	moderately important (0.12)	moderately important (0.15)	moderately important (0.51)	slightly important (0.26)
Fruits	Plum (*Prunus domestica* L.)	17	slightly important (0.05)	moderately important (0.16)	important (0.77)	slightly important (0.33)
Starchy roots	Potato (*Solanum tuberosum* L.)	73	very important (1.84)	very important (2.13)	very important (5.46)	very important (3.14)
Vegetables	Pumpkin (*Cucurbita maxima* Duchesne)	3	important (0.34)	moderately important (0.16)	slightly important (0.26)	slightly important (0.25)
Cereals	Quinoa (*Chenopodium quinoa* Willd.)	9	slightly important (0.02)	slightly important (0.01)	important (0.7)	slightly important (0.24)
Oilcrops	Rapeseed (*Brassica napus* L.)	12	very important (0.61)	very important (0.99)	moderately important (0.55)	important (0.72)
Cereals	Rye (*Secale cereale* L.)	4	moderately important (0.09)	moderately important (0.15)	moderately important (0.44)	slightly important (0.23)
Oilcrops	Safflower (*Carthamus tinctorius* L.)	14	slightly important (0.04)	slightly important (0.02)	important (1.1)	moderately important (0.38)
Cereals	Sorghum (*Sorghum bicolor* (L.) Moench)	17	very important (0.5)	very important (1.02)	important (0.88)	important (0.8)
Oilcrops	Soybean (*Glycine max* (L.) Merr.)	5	very important (3.34)	very important (3.29)	slightly important (0.26)	very important (2.3)
Vegetables	Spinach (*Spinacia oleracea* L.)	2	important (0.34)	moderately important (0.17)	slightly important (0.22)	slightly important (0.24)
Fruits	Strawberry (*Fragaria x ananassa* (Duchesne ex Weston) Duchesne ex Rozier)	17	slightly important (0.05)	slightly important (0.14)	very important (1.64)	important (0.61)
Sugarcrop	Sugar beet (*Beta vulgaris* L.)	14	very important (0.59)	important (0.71)	important (1.1)	important (0.8)
Sugarcrop	Sugarcane (*Saccharum officinarum* L.)	11	very important (1.36)	very important (4.66)	important (0.84)	very important (2.29)
Oilcrops	Sunflower (*Helianthus annuus* L.)	36	very important (1.2)	important (0.62)	very important (2.19)	very important (1.34)

Table 7.2 – continued from previous page

Crop type	Crop name (scientific name)	No. of taxa	Crop value dimensions			
			Contributions to human diets (score)	**Importance to agricultural systems (score)**	**Breeding value (score)**	**Overall value (score)**
Starchy roots	Sweetpotato (*Ipomoea batatas* (L.) Lam.)	14	important (0.47)	important (0.73)	important (0.66)	important (0.62)
Vegetables	Tomato (*Solanum lycopersicum* L.)	12	very important (0.92)	very important (1.42)	important (0.99)	very important (1.11)
Vegetables	Turnip (*Brassica rapa* L.)	8	important (0.34)	moderately important (0.24)	moderately important (0.59)	moderately important (0.39)
Pulses	Urd bean (*Vigna mungo* (L.) Hepper)	21	moderately important (0.09)	slightly important (0.09)	important (0.95)	moderately important (0.38)
Pulses	Vetch (*Vicia sativa* L.)	9	-	slightly important (0.02)	important (0.84)	moderately important (0.43)
Starchy roots	Water yam (*Dioscorea alata* L.)	6	moderately important (0.25)	important (0.6)	moderately important (0.48)	moderately important (0.44)
Fruits	Watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai)	6	slightly important (0.05)	important (0.75)	moderately important (0.4)	moderately important (0.4)
Cereals	Wheat (*Triticum aestivum* L.)	42	very important (8.03)	very important (6.7)	very important (3.31)	very important (6.01)
Starchy roots	White Guinea yam (*Dioscorea rotundata* Poir.)	4	moderately important (0.25)	important (0.6)	moderately important (0.44)	moderately important (0.43)
Starchy roots	Yautia (*Xanthosoma sagittifolium* (L.) Schott)	1	slightly important (0.04)	slightly important (0.01)	slightly important (0.11)	slightly important (0.05)
Vegetables	Zucchini (*Cucurbita pepo* L.)	8	important (0.34)	moderately important (0.16)	important (0.7)	moderately important (0.4)

Crops globally important for their contributions to agricultural production systems include cereals such as barley, rice and sorghum; forages such as alfalfa; fruits such as grapes, banana and plantain; oil crops such as mustard and black mustard; pulses such as lima and common bean; and vegetables such as tomato. The wild relative taxa richness of some of these crops is concentrated in the Mediterranean basin, the Middle East, Central and South America and parts of Central Asia (Figure 7.2a). Sunflower, sweetpotato, mango and yams are among other crops of global importance for agricultural production systems (i.e., "important" category). The distributions of their wild relatives are spread over the world, displaying subtle richness hotspots in Eastern USA and the Malay Archipelago (Figure 7.2b). Moreover, crops of a more regional importance to agricultural production systems include oat, chickpea, carrot, lettuce, pea, rye and spinach (i.e., "moderately important" category). The wild relatives of these crops are distributed in temperate and tropical regions, with notable richness spots in the Iberian Peninsula and part of the Middle East (Figure 7.2c). Strawberry, pigeonpea, almond and apricot are among the crops important for the agricultural production systems of specific regions and/or countries (i.e., "slightly important" category). Their wild relative taxa are accumulated in Western USA, the Iberian Peninsula, the Middle East, South and East Asia (Figure 7.2d).

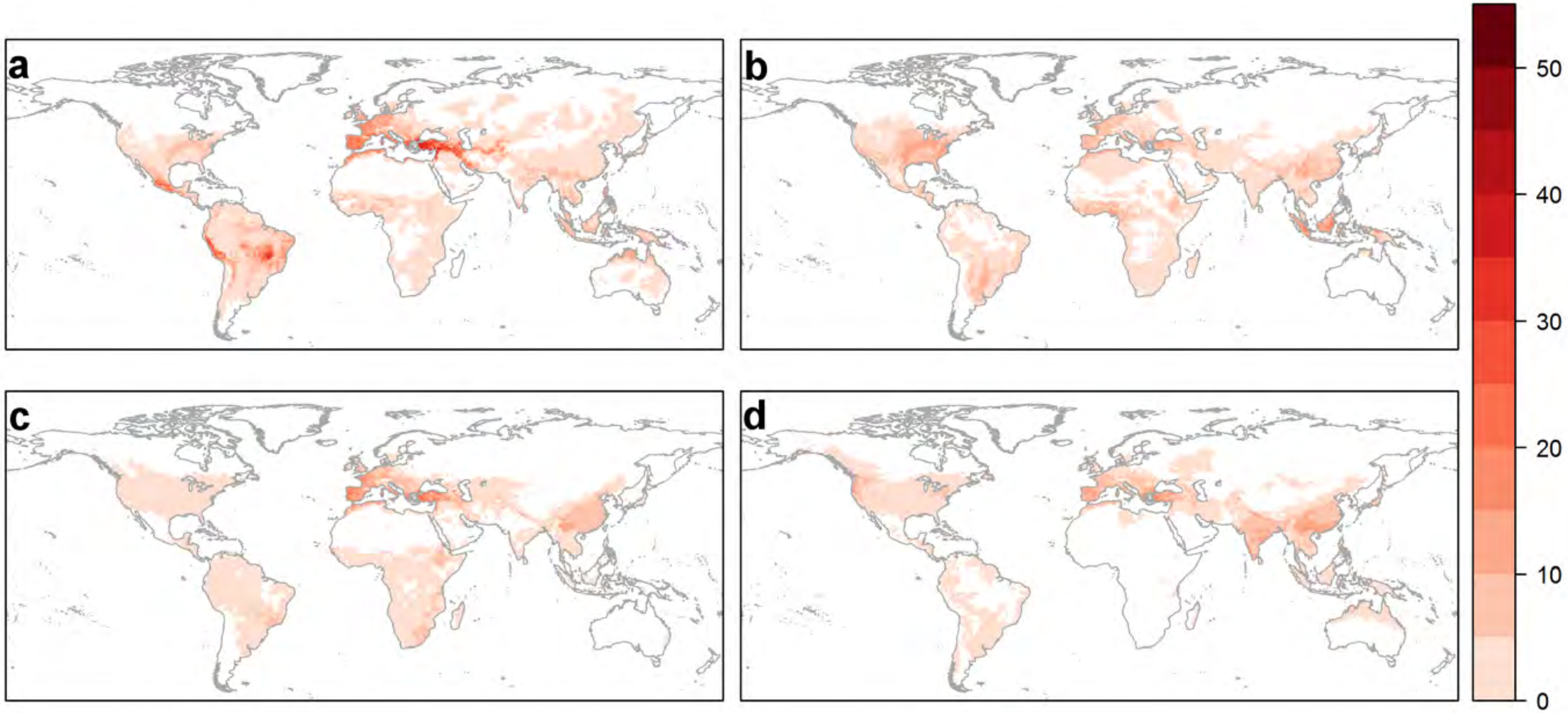


Figure 7.2: Richness maps of crop wild relatives grouped by associated crops' contributions to agricultural productive systems. **a.** Richness map of the wild relatives of crops in the "very important" category according to their contributions to agricultural productive systems; **b.** Richness map of the wild relatives of crops classified in the "important" category due to their contributions to agricultural productive systems; **c.** Richness map of the wild relatives of crops classified in the "moderately important" category for their contributions to agricultural productive systems; **d.** Richness map of the wild relatives of crops in the "slightly important" category for their contributions to agricultural productive systems. Darker red colours indicate high overlap (i.e., high richness) of crop wild relative taxa.

Cereals such as barley, maize and sorghum; fruits such as banana and plantain; oil crops such as groundnut, sunflower, mustards, rapeseed and soybean; sugar crops such as sugar beet and sugarcane; and vegetables such as onions and tomatoes were ranked as very important crops for their contributions of macronutrients to diets. The crop wild relatives' richness spots of some of these crops are found across the Americas, Southern Europe, and the Middle East (Figure 7.3a). Minor cereals such as millets, and vegetables such as asparagus, cabbage, lettuce, carrot and spinach are found in the "important" category for their contributions of calories, fat and protein to global diets. Their wild relatives are distributed in all inhabited continents, and the hotspots of richness are found in the Mediterranean basin (Figure 7.3b). Yams and pulses such as bambara groundnut, pigeonpea, chickpea are among the crops ranked in the "moderately important" category for their contributions of macronutrients to global diets. The wild relatives of these crops are mainly distributed in the tropics, with no remarkable richness hotspots (Figure 7.3c). Fruits, including almond, apricot, cherry, grapefruit, mango and peach are found in the first quartile of crops contributing macronutrients to global diets (i.e., "slightly important category"). The wild relatives of the crops in the "slightly important" category are found in the Americas, Europe, Southern Africa, and Asia, with richness hotspots in the Middle East, East Asia and the Malay Archipelago (Figure 7.3d).

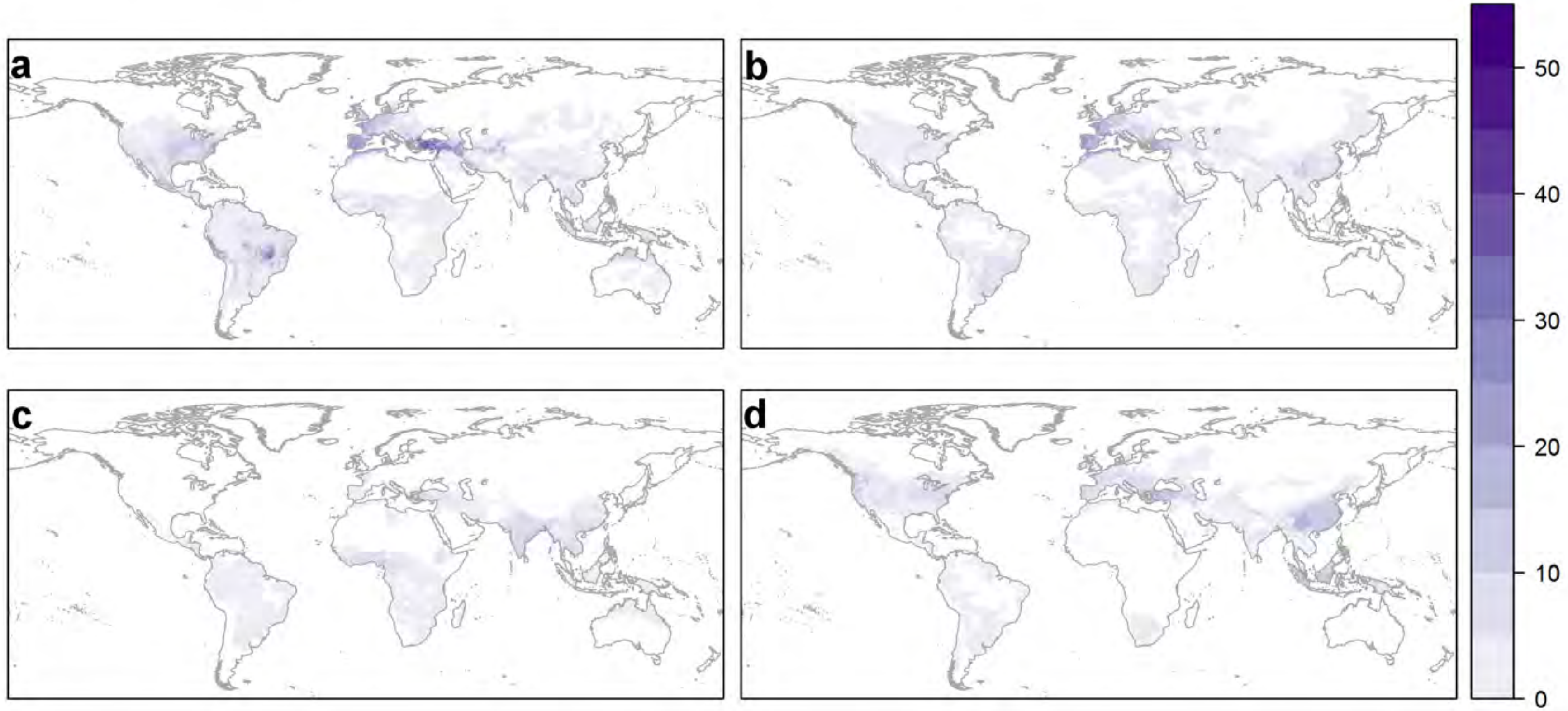


Figure 7.3: Richness maps of crop wild relatives grouped by associated crops macronutrient contributions to human diets. **a.** Richness map of the wild relatives of crops classified as very important according to their global contributions to human diets; **b.** Richness map of the wild relatives of crops classified in the “important” category due to their contributions to human diets; **c.** Richness map of the wild relatives of crops classified in the “moderately important” category for their contributions to human diets; **d.** Richness map of the wild relatives of crops in the “slightly important” category for their contributions to human diets. Darker purple colours indicate high overlap (i.e., high richness) of crop wild relative taxa.

Forages such as alfalfa; fruits such as almond, apricot, breadfruit, cherry, apple, mango, peach, pear and strawberry; oil crops such as cottonseed as sunflower; pulses such as cowpea; and vegetables such as cabbage, carrot and eggplant are among the crop gene pools highly likely to be used in plant breeding (i.e., "very important" category). Conspicuous spots with more than 15 overlapping taxa are found in Central and South America, Eastern USA, the Mediterranean basin, the Middle East, East and South East Asia (Figure 7.4a). Asparagus, oat, sugar beet and lettuce are among the crop gene pools in the "important" category of the likelihood of use in plant breeding metric. The crop gene pools in this category contain thirteen wild relative taxa in average, including wild relative taxa that can present challenges at the moment of being used in breeding (i.e., tertiary gene pool). The wild relatives of these crops are distributed worldwide with hotspots of richness in the Middle East and the Iberian Peninsula (Figure 7.4b). Millets, yams, citrus and bananas are among the crop gene pools in the "moderately" and "slightly" important categories. These crops present narrow gene pools (five wild relative taxa in average) and an elevated proportion of wild relative taxa of the tertiary gene pool. The wild relatives of these crop gene pools are evenly distributed in tropical and sub-tropical regions with no remarkable richness hotspots (Figures 7.4c and 7.4d).

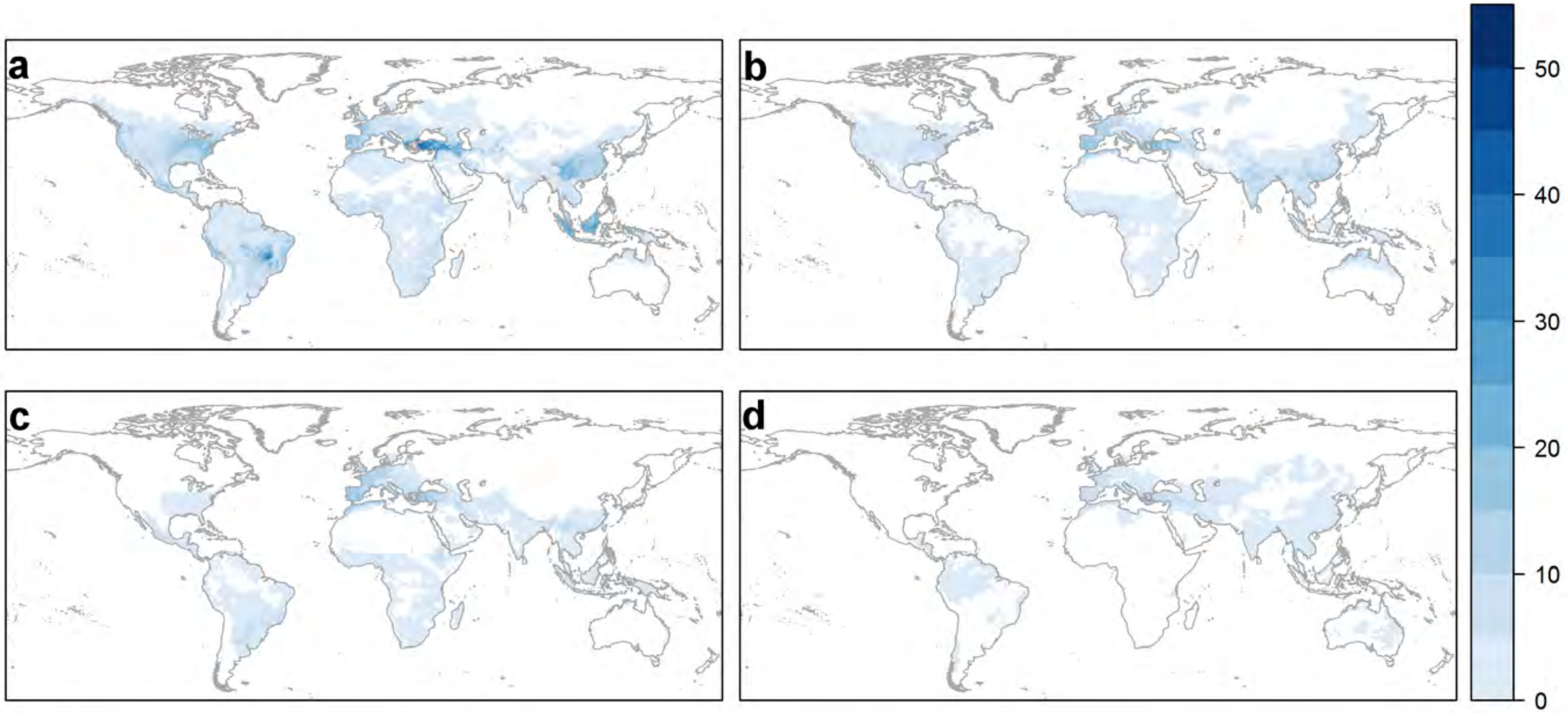


Figure 7.4: Richness maps of crop wild relatives grouped by their crop gene pool likeliness of being used in plant breeding. **a.** Richness map of the wild relatives of crops classified in the “very important” category according to their likeliness of being used in plant breeding; **b.** Richness map of the wild relatives of crops classified in the “important” category due to their likeliness of being used in plant breeding; **c.** Richness map of the wild relatives of crops classified in the “moderately important” category for their likeliness of being used in plant breeding; **d.** Richness map of the wild relatives of crops in the “slightly important” category for their likeliness of being used in plant breeding. Darker blue colours indicate high overlap (i.e., high richness) of crop wild relative taxa.

7.4.2 Complementarity between prioritization scores

The collecting priority score and the crop value dimensions calculated in this study were found to be weakly correlated (Figure 7.5). The crop value dimensions here examined could serve as complementary rules to further prioritize crop gene pools for conservation as each of them provide unique information, not yet represented by the Final Priority Score (FPS; Chapter 5). High correlation coefficients were found between two crop value dimensions: contribution to human diets and importance to agricultural systems (ρ = 0.89). This could be explained by the high dependence of global food supply systems on few crops (Khoury *et al.*, 2014). Both dimensions are further included because of the information each dimension provides.

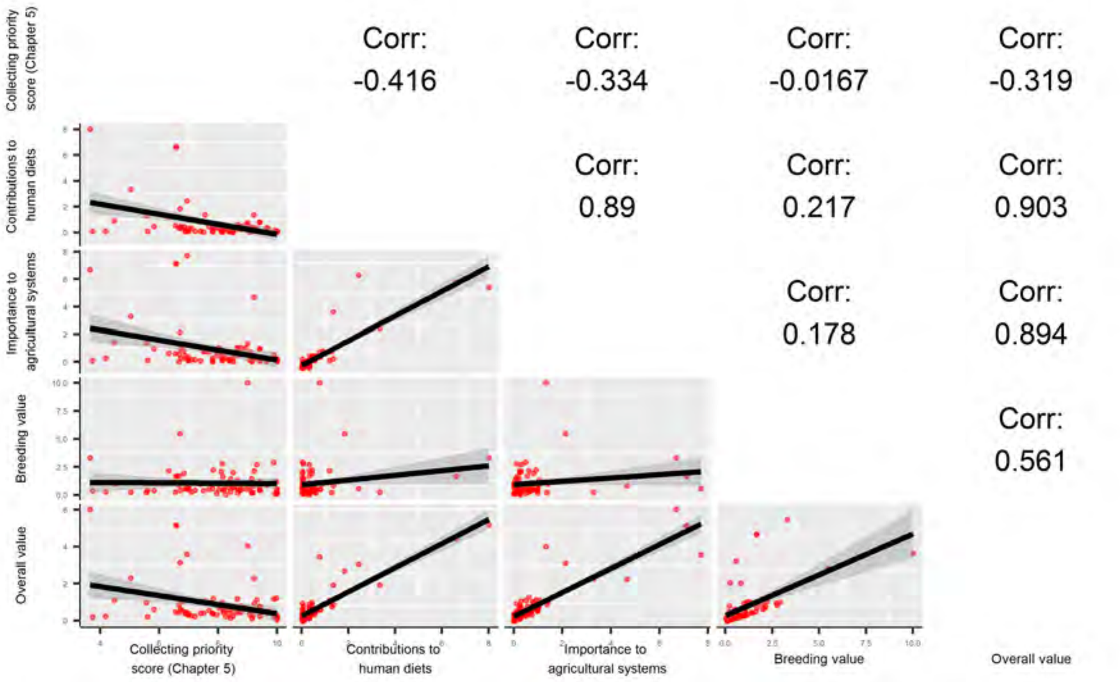


Figure 7.5: Correlation matrix of crop value dimensions and collecting priority score (Final Priority Score in Chapter 5). Each red dot represents one crop gene pool.

Fifty-eight crop gene pools were ranked as high priorities requiring urgent collecting and further *ex situ* conservation due to their low representativeness in genebanks (Chapter 5). If further parameters for refining the conservation priorities are applied, cassava would rank as being of highest priority for conservation given its importance to agricultural production systems (ranked in the “very important” category), its contributions to diets (“very important” category) and likelihood of being used in plant breeding (again, “very important” category). A combination of two, three or four prioritization parameters can help guiding conservation efforts, while investing scarce resources more efficiently. Deciding which is the most appropriate combination of parameters will depend directly on the objectives’ and scope of the conservation effort. Here a non-exhaustive list of priority crop gene pools after applying further refinements of conservation priorities by combining two parameters:

- High priorities for *ex situ* conservation + very important crops for their contributions to diets: Sugar beet, cassava, groundnut, banana and plantain, mustard, black mustard, rapeseed, sorghum and sugarcane.
- High priorities for *ex situ* conservation + very important crops for global agricultural production systems: Cassava, grape, alfalfa, banana and plantain, mustard, black mustard, rapeseed, sorghum and sugarcane,
- High priorities for *ex situ* conservation + high breeding value: Almond, apricot, breadfruit, cassava, cherry, cowpea, carrot, eggplant, apple, mango, alfalfa, peach, pear and strawberry.
- Very important crops for their contributions to diets and global agricultural production systems: Cassava, barley, maize, banana and plantain, mustard, black mustard, potato, rapeseed, African and Asian rice, sorghum, soybean, sugarcane, tomato and wheat.
- Very important crops for their contributions to diets and high breeding value: Cassava, sunflower, potato, African and Asian rice, and wheat.
- Very important crops for their contributions to global agricultural production systems and high breeding value: Cassava, alfalfa, potato, African and Asian rice, and wheat.

7.5 Discussion

The methods and results herein presented are a proposal to estimate the value of crop wild relatives using multiple parameters, and are expected to help shaping future conservation efforts of crop wild relatives. The importance categories used to classify each crop value dimension permit the identification and grouping of crops according to their global or regional importance, and the likelihood of crop gene pools to be used in plant breeding programs given the easiness of producing fertile offspring. Crop wild relative taxa richness hotspots were identified for crops globally important for food and agriculture (very important and important categories), and highly likely to be incorporated in plant breeding programs (very important and important categories). In the same way, the analysis presents the richness patterns of crop wild relatives associated to crops of more regional importance (moderately and slightly important categories), and crop gene pools that can result more challenging to be used in plant breeding (moderately and slightly important categories).

The crop value dimensions estimated herein can be combined to prioritize crop gene pools for conservation, according to the objectives and scope of the conservation effort. For instance, chapter 5 presents the prioritization of crop gene pools based on their extent of genetic diversity represented in genebanks. The crop value dimensions estimated herein are complementary to this collecting prioritization score and can be used to further refine *ex situ* conservation actions, resulting in a short list of crop gene pools that require immediate attention.

South America, the Mediterranean basin, the Middle East, East and Southeast Asia emerge as richness hotspots of wild relative taxa of crops of global and regional importance for their contributions to agricultural production systems and/or global diets. Furthermore, these hotspots also harbor crop gene pools likely to be used in plant breeding. These richness hotspots are candidate sites for establishing *in situ* reserves, as they hold an important wealth of diversity beneficial for multiple crops. Detecting and complementing ongoing conservation strategies focused on conserving habitats, rather than single species is highly desirable, as it is a mean to secure long-term conservation of crop wild relatives and other beneficial organisms like pollinators and associated microbiota. Moreover, these regions were also identified as collecting priorities to improve the current representativeness of crop wild relatives held in genebanks (Chapter 5). *Ex situ* conservation can also serve as backup of existing genetic diversity, and is a strategy that eases access to genetic resources to plant breeders and researchers. Particular attention is required to the richness hotspots located in the Middle East, South and East Asia, as these areas have been targeted as being vulnerable to climate change due to their projected climatic instability and low degree of vegetation intactness (Watson *et al.*, 2013).

CHAPTER 8

CONCLUSIONS

The research presented in this thesis provides a global overview of the conservation concerns of crop wild relatives. Species occurrence data, species distribution models and *ex situ* gap analyses were used to understand the extent of the potential diversity that is currently conserved in genebanks, and therefore setting priorities for future collecting. Geographic regions where crop wild relatives are distributed and concentrated were also identified and used to prioritize regions where conservation efforts need to be prioritized. Climate change projections were used to assess the potential impact of climate change on the distributions of crop wild relatives, and to identify crop gene pools and crop wild relative taxa expected to be largely affected by climate change. And a set of complementary dimensions was used to estimate the value of crop gene pools as a means to further refine conservation priorities at the global and regional scale. The results presented in this thesis constitute the first quantitative assessment of the extent of potential diversity of crop wild relatives conserved in genebanks for two solanaceous crop gene pools (Chapters 3 and 4) and 81 crop gene pools (Chapter 5), and the likely impacts of climate change in several crop wild relatives (Chapter 6).

8.1 Main findings and implications

The analyses presented in this study had led to the following conclusions:

1. The growing availability of species occurrence data in digital and reusable formats, to-

gether with global dataset of meaningful bioclimatic variables and species modelling algorithms were fundamental to pursue this study. Filling the information gaps of occurrence records and addressing some of the quality issues (e.g., taxonomic identification of occurrence records) are instrumental to improve the distributions of crop wild relatives, produce more accurate representations of the distributions of crop wild relatives, and in consequence the improvement of the analyses that can be derived and used to guide conservation efforts (Chapter 2).

2. The crop wild relatives are poorly represented in genebanks (Chapter 5). Improving the representativeness of their potential diversity in genebanks requires large efforts for collecting propagules from the natural habitats of crop wild relatives. This in turn contributes to secure the long-term conservation of crop wild relatives, to improve the availability of these important genetic resources to the global community, and to achieve global commitments related to food security and biodiversity conservation (i.e., SDG's and Aichi Biodiversity Targets (Secretariat of the CBD, 2010)). Further investments are therefore required to achieve a more comprehensive conservation of crop wild relatives. Likewise, political challenges need to be addressed to facilitate the collection, conservation and use of crop wild relatives globally (Esquinas-Alcázar, 2005).

3. The *ex situ* gap analyses results presented here can be used as a baseline of the current state of the potential diversity of 1,076 crop wild relatives conserved in genebanks. This methodology has been improved to analyze different crop gene pools, and therefore can be applied routinely to assess the progress towards the comprehensive conservation of these genetic resources.

4. The wild relatives of apple, cassava, grape, groundnut, sorghum, sugar beet, sugarcane and sweetpotato were identified as high priorities for conservation due to their low potential diversity currently represented in genebanks (Chapter 5), contributions of their associated crops to global human diets and agricultural production systems, and likelihood of being used in plant breeding (Chapter 7). Hotspots for further collecting these wild rel-

atives include: South and North America, East and Southeast Asia, tropical Africa, and the Near East.

5. For the particular case of potatoes and eggplants (Chapters 3 and 4), 43.8% and 76.9% of their wild relatives were considered of urgent conservation due to their low representativeness in genebanks. Field observations of potato and eggplant wild relatives, and preliminary IUCN Red List assessments (only for eggplant wild relatives) identified particular species currently threatened, and therefore urgent to conserve. Further improvements in the taxonomic identification, specially of genebank accessions, were detected as limitations in both case studies. The definition of plants' species and their taxonomic identities are dynamic, and these changes are not immediately reflected in genebank collections. Alternative mechanisms to identify species (e.g., DNA barcoding) may help to overcome this bottleneck (Jarret, 2008; Girma *et al.*, 2016). However, complex taxonomic groups, such as potato, require multi-dimensional approaches to define the species' boundaries (Spooner, 2009).

6. The evaluation of the potential impacts of climate change on the distributions of crop wild relative revealed an average of 20.8% of suitable area lost due to climate change. One in ten crop wild relative species are under the risk of losing more than 50% of their current suitable areas. Finger millet, cowpea and potato were identified as the gene pools (i.e., group of wild relatives of a crop) more negatively impacted, whereas, the crop gene pools of rice and pigeonpea are likely to be less affected by the increasing temperatures and more variable precipitation of climate change.

8.2 Limitations

The following limitations were found during the preparation of this study. First, despite the growing availability of species occurrence records, many taxonomic groups and geographic regions remain unevenly represented. Lack of sufficient data prevented the analysis of crop gene pools important for their contributions to agriculture such as tea, coffee and avocado. Strategies to close this information gap include the establishment of collaborations with botanists, that may have accumulated large amounts of field observations and botanical specimens on the wild relative species of globally important crops. Likewise, global initiatives to digitize living collections (e.g., herbarium specimens) and germplasm collections may reveal adequate quantities of data for performing analyses as those presented in this study.

Species distribution modelling is an increasingly useful tool for conservation studies (Johnson and Gillingham, 2005; Guisan *et al.*, 2013). The settings and parameterization utilized in this thesis are considered adequate for the scale of this study (i.e., global scale) (Phillips *et al.*, 2006). More localized analysis can be produced by selecting the adequate parameters for the species (or group of species) under analysis, and the geographic extent of the analysis. This includes determining the environmental drivers of the distribution of species, the background extent to be used with the model, the dataset to be used for training and testing the species distribution model, and the modelling algorithm to be used.

The gap analysis is an approach that helps estimating the potential diversity encompassed in crop wild relatives that is currently conserved and available via genebanks, and may be used as an alternative given the general lack of molecular data available for this species. Moreover, targeting geographic hotspots for further collecting and collecting are mandatory activities to acquire sufficient samples to perform sound molecular analyses. Given the more affordable costs of molecular analyses (Sboner *et al.*, 2011), it is expected that further refinements of the conservation priorities here presented can be performed (McCouch *et al.*, 2013). Once molecular data on crop wild relatives is produced, field collection planning can be complemented by using such data to understand the genetic composition and structure of plant populations, therefore aiding to prioritize those populations (or sub-populations) containing unique alleles and/or alleles of interest for plant breeding (Camadro, 2012).

The preliminary IUCN Red List assessment is largely based in geographic assessments (Chapter 4 ; Bachman *et al.*, 2011). This approach utilizes georeferenced occurrence data to categorize crop wild relatives in threat categories. Filling the gaps of occurrence data can help improving the availability of records within a species' range. This, in combination with population data will definitely help producing formal assessments of the level of threat of crop wild relatives, and therefore conduct the priorities setting for the conservation of such species in their natural habitats (e.g., Rodríguez Delgado *et al.*, 2011).

As per the assessment of the climate change impacts on the distribution of crop wild relatives (Chapter 6), only one stringent representative concentration pathway (RCP 4.5) was considered in this study. Additional analyses, including more than one representative concentration pathway (e.g., best and worse case scenarios), and several time slices enable the production of comparative scenarios of the impact that crop wild relatives may face if actions to reduce the green-house gas emissions are (or are not) implemented globally (e.g., Warren *et al.*, 2013).

Chapter 7 uses a set of dimensions to refine global conservation priorities of crop wild relatives. The dimension that represents the contributions of crops to global diets is exclusively based on macronutrients such as proteins, fats and calories (see Food Balance Sheets in `http://faostat3.fao.org/`). In consequence, the classification of highly important crops to global diets is biased towards crops rich in starchs, carbohydrates, oils, and proteins, such as cereals, oilcrops, starchy roots and sugarcrops. Dissagregated data on the contributions of micronutrients and fiber to diets may help to highlight the role of fruits and vegetables to healthy diets.

8.3 Future work

The work presented in this thesis has enabled the author to identify the following opportunities for future work. First, the work presented with regards to the *ex situ* gap analysis of crop wild relatives represents the first exhaustive assessment of the extent of potential diversity of 1,076 crop wild relatives conserved in genebanks and accessible to the global community. The growing affordability of molecular analysis can be used to study the genetic structure of crop wild relative species' populations, leading to the identification of populations that are not yet represented in genebanks and therefore need to be collected (e.g., Hoffmann *et al.*, 2003).

Likewise, species distribution models projected onto past, present and future climates can be used to detect refugia areas associated to rich genetic diversity (Thomas *et al.*, 2012; Russell *et al.*, 2014).

The *ex situ* gap analysis is based in three metrics that are used as proxies of the potential diversity of crop wild relative species. Quantifying the contributions of geographical and ecological distances to the genetic patterns of diversity is a clear area for future work. This will help understanding to what extent the gap analysis metrics are able to capture and represent the potential genetic diversity possessed by crop wild relatives.

The final priority score estimated with the *ex situ* gap analysis can be adapted to be used as an indicator to quantify the progress towards the *ex situ* conservation of crop wild relatives. This will need to re-calculate the gap analysis, using refreshed germplasm accessions data on a routine basis. This can be useful to track the achievement of the Sustainable Development Goal 2 (Target 5), and the Aichi Biodiversity Target 13) with regards to the conservation of wild plant germplasm.

With regards to threats assessments, additional studies can help to estimate the extent to which crop wild relatives are exposed to increasing human pressures (Geldmann *et al.*, 2014). And to identify the crop wild relatives that may occur in ecosystems sensible to climate variability (Seddon *et al.*, 2016), and climatic change (Watson *et al.*, 2013).

In terms of conservation actions, much work is needed in order to collect, conserve and use crop wild relatives. International agreements governing the collecting and exchange of plant genetic resources play a critical role to enable and promote the use of these genetic resources (López Noriega *et al.*, 2013; Heisey and Rubenstein, 2015). Swiftly implementation and adoption of such agreements is urgently needed, specially for the countries interested in strengthening their agricultural productive systems.

Additional efforts to establish collaborations and alliances with other conservation beyond the plant genetic resources are needed if the comprehensive conservation of crop wild relatives is desired. For instance, botanical gardens can also play an important role in the conservation of crop wild relatives, particularly for those that due to their seed storage behavior are better maintained in the field. This said, enhancing the visibility of accessions held in botanical gardens through widely known portals to plant breeders and plant genetic resources community (see Genesys; `https://www.genesys-pgr.org`) can be the first step towards the establishment of such collaborations. Likewise, the implementation and adoption of the protocols such as the Multi-crop passport descriptors (Alercia *et al.*, 2012), and material transfer agreements will be needed to facilitate access to such wild relative accessions.

LIST OF REFERENCES

Abberton, M., Batley, J., Bentley, A., Bryant, J., Cai, H., Cockram, J., Costa de Oliveira, A., Cseke, L. J., Dempewolf, H., De Pace, C., Edwards, D., Gepts, P., Greenland, A., Hall, A. E., Henry, R., Hori, K., Howe, G. T., Hughes, S., Humphreys, M., Lightfoot, D., Marshall, A., Mayes, S., Nguyen, H. T., Ogbonnaya, F. C., Ortiz, R., Paterson, A. H., Tuberosa, R., Valliyodan, B., Varshney, R. K. and Yano, M. (2015). Global agricultural intensification during climate change: a role for genomics, *Plant Biotechnology Journal* pp. 1–4.

Afonin, A. and Greene, S. L. (1999). Germplasm collecting using modern modern geographic information technologies: directions explored by the N.I. Vavilov Institute of Plant Industry, *in* S. L. Greene and L. Guarino (eds), *Linking Genetic Resources and Geography: Emerging Strategies for Conserving and Using Crop Biodiversity*, number 27, Crop Science Society of America and American Society of Agronomy, Madison, WI, chapter 6, pp. 75–85.

Ahammad, H., Clark, H., Dong, H., Elsiddig, E. A., Haberl, H., Harper, R., House, J., Jafari, M., Masera, O., Mbow, C., Ravindranath, N. H., Rice, C. W., Abad, C. R., Romanovskaya, A., Sperling, F. and Tubiello, F. N. (2014). Agriculture, Forestry and Other Land Use (AFOLU), *Climate Change 2014: Mitigation of Climate Change. Working Group III Contribution to the IPCC 5th Assessment Report /Changes to the underlying Scientific/Technical Assessment*, chapter 11, pp. 1–179.

Ahmadi, H. and Bringhurst, R. S. (1992). Breeding Strawberries at the Decaploid Level, *Journal of the American Society for Horticultural Science* **117**(5): 856–862.

Ahuja, S., Mukhopadhya, M., Singh, C. and Ahuja, S. (1987). Effects of infestation of eggplant (*Solanum melongena*) with root knot nematode (*Meloidogyne incognita*) on the oxidative enzymes and cell wall constituents in their roots, *Capsicum Newsletter* **6**: 98–99.

Ainsworth, E. a., Yendrek, C. R., Sitch, S., Collins, W. J. and Emberson, L. D. (2012). The effects of tropospheric ozone on net primary productivity and implications for climate change, *Annual review of plant biology* **63**: 637–61.

Alercia, A., Diulgheroff, S. and Mackay, M. (2012). FAO/Bioversity Multi-crop passport descriptors V.2, p. 11.

Alexander, J. M., Diez, J. M. and Levine, J. M. (2015). Novel competitors shape species' responses to climate change, *Nature* **525**: 515–518.

Ames, M. and Spooner, D. M. (2010). Phylogeny of *Solanum* series Piurana and related species in *Solanum* section Petota based on five conserved ortholog sequences, *Taxon* **59**(4): 1091–1101.

Anthony, F., Bertrand, B., Etienne, H. and Lashermes, P. (2011). *Coffea* and *Psilanthus*, *in* C. Kole (ed.), *Wild Crop Relatives: Genomic and Breeding Resources*, Springer Berlin Heidelberg, Berlin, Heidelberg, chapter 3, pp. 41–61.

Araújo, M. B. and New, M. (2007). Ensemble forecasting of species distributions, *Trends in ecology & evolution* **22**(1): 42–7.

Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Prasad, P. V. V., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Challinor, A. J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurralde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A.-K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J. E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P. J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z. and Zhu, Y. (2015). Rising temperatures reduce global wheat production, *Nature Climate Change* **5**(2): 143–147.

Aubriot, X., Singh, P. and Knapp, S. (2016). Tropical Asian species show the Old World clade of "spiny solanums" (subgenus *Leptostemonum* Bitter, the Leptostemonum Clade: Solanaceae) is not monophyletic, *Botanical Journal of the Linnean Society* **181**: 199–223 .

Bachman, S. P., Moat, J., Hill, A., de la Torre, J. and Scott, B. (2011). Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool, *ZooKeys* **150**: 117–126.

Beebe, S., Ramirez, J., Jarvis, A., Rao, I. M., Mosquera, G., Bueno, J. M. and Blair, M. W. (2011). Genetic improvement of common beans and the challenges of climate change, *in* S. S. Yadav, R. Redden, J. L. Hatfield, H. Lotze-Campen and A. J. W. Hall (eds), *Crop Adaptation to Climate Change*, Wiley-Blackwell, chapter Chapter 16, pp. 356–369.

Berdegué, J. A. and Fuentealba, R. (2011). Latin America: The State of Smallholders in Agriculture, *IFAD conference on new directions for smallholder agriculture*, Rome.

Bilz, M., Kell, S. P., Maxted, N. and Lansdown, R. V. (2011). *European Red List of Vascular Plants*, Publications Office of the European Union, Luxembourg.

Bivand, R., Keitt, T. and Rowlingson, B. (2014). *rgdal: Bindings for the Geospatial Data Abstraction Library*.
URL: *http://cran.r-project.org/package=rgdal*

Bivand, R. and Lewin-Koh, N. (2014). *maptools: Tools for reading and handling spatial objects*.
URL: *http://cran.r-project.org/package=maptools*

Bivand, R. S., Pebesma, E. and Gomez-Rubio, V. (2013). *Applied Spatial Data Analysis with R.*, second ed edn, Springer-Verlag, New York.

Bletsos, F. and Roupakias, D. (1998). Interspecific hybrids between three eggplant (*Solanum melongena* L.) cultivars and two wild species (*Solanum torvum* Sw. and *Solanum sisymbriifolium* Lam.), *Plant Breeding* **164**: 159–164.

Bodirsky, B. L., Popp, A., Lotze-Campen, H., Dietrich, J. P., Rolinski, S., Weindl, I., Schmitz, C., Müller, C., Bonsch, M., Humpenöder, F., Biewald, A. and Stevanovic, M. (2014). Reactive nitrogen requirements to feed the world in 2050 and potential to mitigate nitrogen pollution, *Nature communications* **5**(3858).

Bohs, L. (2005). Major clades in *Solanum* based on ndhF sequence data, *in* R. C. Keating, V. C. Hollowell and T. B. Croat (eds), *A festschrift for William G. D'Arcy: the legacy of a taxonomist*, Missouri Botanical Garden Press, St Louis, Missouri, USA, pp. 27–49.

Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., Matasci, N., Narro, M. L., Piel, W. H., McKay, S. J., Lowry, S., Freeland, C., Peet, R. K. and Enquist, B. J. (2013). The taxonomic name resolution service: an online tool for automated standardization of plant names, *BMC Bioinformatics* **14**(1): 16.

Bradshaw, J. E., Bryan, G. J. and Ramsay, G. (2006). Genetic Resources (Including Wild and Cultivated *Solanum* Species) and Progress in their Utilisation in Potato Breeding, *Potato Research* **49**(1): 49–65.

Bradshaw, J. E. and Ramsay, G. (2005). Utilisation of the Commonwealth Potato Collection in potato breeding, *Euphytica* **146**(1-2): 9–19.

Brar, D. and Khush, G. (1997). Alien introgression in rice, *Plant Molecular Biology* **35**(1-2): 35–47.

Brehm, J. M., Maxted, N., Martins-Loução, M. A. and Ford-Lloyd, B. V. (2010). New approaches for establishing conservation priorities for socio-economically important plant species, *Biodiversity and Conservation* **19**(9): 2715–2740.

Brummitt, N. and Bachman, S. (2010). Plants under pressure - a global assessment. The first report of the IUCN Sampled Red List Index for Plants, *Technical report*, Royal Botanic Gardens, Kew, UK, London.

Brummitt, N., Bachman, S. P., Aletrari, E., Chadburn, H., Griffiths-Lee, J., Lutz, M., Moat, J., Rivers, M. C., Syfert, M. M. and Lughadha, E. M. N. (2015a). The Sampled Red List Index for Plants, phase II: ground-truthing specimen-based conservation assessments, *Philosophical Transactions of the Royal Society B* **370**(1): 1.

Brummitt, N. A., Bachman, S. P., Griffiths-Lee, J., Lutz, M., Moat, J. F., Farjon, A., Donaldson, J. S., Hilton-Taylor, C., Meagher, T. R., Albuquerque, S., Aletrari, E., Andrews, A. K., Atchison, G., Baloch, E., Barlozzini, B., Brunazzi, A., Carretero, J., Celesti, M., Chadburn, H., Cianfoni, E., Cockel, C., Coldwell, V., Concetti, B., Contu, S., Crook, V., Dyson, P., Gardiner, L., Ghanim, N., Greene, H., Groom, A., Harker, R., Hopkins, D., Khela, S., Lakeman-Fraser, P., Lindon, H., Lockwood, H., Loftus, C., Lombrici, D., Lopez-Poveda, L., Lyon, J., Malcolm-Tompkins, P., McGregor, K., Moreno, L., Murray, L., Nazar, K., Power, E., Quiton Tuijtelaars, M., Salter, R., Segrott, R., Thacker, H., Thomas, L. J., Tingvoll, S., Watkinson, G., Wojtaszekova, K. and Nic Lughadha, E. M. (2015b). Green plants in the red: a baseline global assessment for the IUCN Sampled Red List Index for plants, *Plos One* **10**(8): e0135152.

Cadima, X., van Zonneveld, M., Scheldeman, X., Castañeda, N., Patiño, F., Beltran, M., Van Damme, P. and Beltrán, M. (2014). Endemic wild potato (*Solanum* spp.) biodiversity status in Bolivia: Reasons for conservation concerns, *Journal for Nature Conservation* **22**(2): 113–131.

Cai, D., Rodríguez, F., Teng, Y., Ané, C., Bonierbale, M., Mueller, L. A. and Spooner, D. M. (2012). Single copy nuclear gene analysis of polyploidy in wild potatoes (*Solanum* section Petota), *BMC evolutionary biology* **12**(1): 70.

Camadro, E. L. (2012). Relevance of the genetic structure of natural populations, and sampling and classification approaches for conservation and use of wild crop relatives: potato as an example, *Botany* **11**: 1065–1072.

Camadro, E. L. (2014). Is the genetic integrity of natural plant populations *ex situ* preserved with the current sampling, conservation and regeneration approaches?, *Journal of Basic and Applied Genetics* **25**(1): 41–44.

Camadro, E. L., Carputo, D. and Peloquin, S. J. (2004). Substitutes for genome differentiation in tuber-bearing *Solanum*: interspecific pollen-pistil incompatibility, nuclear-cytoplasmic male sterility, and endosperm, *Theoretical and Applied Genetics* **109**(7): 1369 –1379.

Cardi, T., D'Ambrosio, E., Consoli, D., Puite, K. J. and Ramulu, K. S. (1993). Production of somatic hybrids between frost-tolerant *Solanum commersonii* and *S. tuberosum*: characterization of hybrid plants, *TAG. Theoretical and applied genetics.* **87**(1-2): 193–200.

Carputo, D., Cardi, T., Speggiorin, M., Zoina, A. and Frusciante, L. (1997). Resistance to blackleg and tuber soft rot in sexual and somatic interespecific hybrids with different genetic background, *American Potato Journal* **74**: 161–172.

Carputo, D., Speggiorin, M., Garrefa, P., Raio, A. and Monti, L. (1996). Screening for resistance to tuber soft rot and blackleg in diploid *Solanum* species and *S. tuberosum* haploids, *Journal of Genetics and Breeding* **50**: 221–226.

Castañeda-Álvarez, N. P., de Haan, S., Juárez, H., Khoury, C. K., Achicanoy, A., Sosa, C. C., Bernau, V., Salas, A., Heider, B., Simon, R., Maxted, N. and Spooner, D. M. (2015). *Ex Situ* Conservation Priorities for the Wild Relatives of Potato (*Solanum* L . Section Petota), *PLoS ONE* **10**(4): e0122599.

Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., Guarino, L., Harker, R. H., Jarvis, A., Maxted, N., Müller, J. V., Ramirez-Villegas, J., Sosa, C. C., Struik, P. C., Vincent, H. and Toll, J. (2016). Global conservation priorities for crop wild relatives, *Nature Plants*.
URL: *http://dx.doi.org/10.1038/NPLANTS.2016.22*

Castañeda-Álvarez, N. P., Vincent, H. A., Kell, S. P., Eastwood, R. J. and Maxted, N. (2011). Chapter 14: Ecogeographic surveys, *in* L. Guarino, R. V. Ramanatha and E. Goldberg (eds), *Collecting Plant Genetic Diversity:Technical Guidelines - 2011 Update*, Rome, Italy, chapter 14.
URL: *http://cropgenebank.sgrp.cgiar.org/index.php?option=com_content&view=article&id=390&Item*

Cayuela, L., Granzow-de la Cerda, Í., Albuquerque, F. S. and Golicher, D. J. (2012). taxonstand: An R package for species names standardisation in vegetation databases, *Methods in Ecology and Evolution* **3**(6): 1078–1083.

CBD (1992). Convention on Biological Diversity.

CBD (2010). Global Strategy for Plant Conservation (2011-2020).

Ceccarelli, S., Grando, S., Maatougui, M., Michael, M., Slash, M., Haghparast, R., Rahmanian, M., Taheri, A., Al-Yassin, A., Benbelkacem, A., Labdi, M., Mimoun, H. and Nachit, M. (2010). Plant breeding and climate changes, *The Journal of Agricultural Science* **148**(06): 627–637.

CGRFA (2012). *Second global plan of action for plant genetic resources for food and agriculture*, Food and Agriculture Organization of the United Nations, Rome, Italy.

Chae, J.-H., Ha, B.-K., Chung, G., Park, J.-E., Park, E., Ko, J.-M., Shannon, J. G., Song, J. T. and Lee, J.-D. (2015). Identification of environmentally stable wild soybean genotypes with high alpha-linolenic acid concentration, *Crop Science* **55**(4): 1629.

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R. and Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation, *Nature Climate Change* **4**: 287–291.

Chamberlain, S., Ram, K., Barve, V. and Mcglinn, D. (2015). rgbif: Interface to the Global Biodiversity Information Facility 'API'.
URL: *https://github.com/ropensci/rgbif*

Chelliah, S. and Srinivasan, K. (1983). Resistance in bhindi, brinjal and tomato to major insect and mite pests, *Proceedings of the National Seminar of Breeding Crop Plants for Resistance to Pests and Diseases* p. 47.

Chen, I.-c., Hill, J. K., Ohlemüller, R., Roy, D. B. and Thomas, C. D. (2011). Rapid range shifts of species of climate warming, *Science* **333**: 1024–1026.

Collonnier, C., Fock, I., Daunay, M.-C., Servaes, A., Vedel, F., Siljak-Yakovlev, S., Souvannavong, V. and Sihachakr, D. (2003a). Somatic hybrids between *Solanum melongena* and *S. sisymbrifolium*, as a useful source of resistance against bacterial and fungal wilts, *Plant Science* **164**(5): 849–861.

Collonnier, C., Fock, I., Mariska, I., Servaes, A., Vedel, F., Siljak-Yakovlev, S., Souvannavong, V. and Sihachakr, D. (2003b). GISH confirmation of somatic hybrids between *Solanum melongena* and *S. torvum*: assessment of resistance to both fungal and bacterial wilts, *Plant Physiology and Biochemistry* **41**(5): 459–470.

CONABIO (2012). Maíces: Biodiversidad Mexicana.
URL: *http://www.biodiversidad.gob.mx/usos/maices/maiz.html*

Cordell, D., Drangert, J.-O. and White, S. (2009). The story of phosphorus: Global food security and food for thought, *Global Environmental Change* **19**(2): 292–305.

Dale, P. J. (1992). Spread of Engineered Genes to Wild Relatives, *Plant physiology* **100**: 13–15.

Daunay, M. C., Chaput, M. H., Sihachakr, D., Allot, M., Vedel, F. and Ducreux, G. (1993). Production and characterization of fertile somatic hybrids of eggplant (*Solanum melongena* L.) with *Solanum aethiopicum* L., *TAG. Theoretical and applied genetics.* **85**(6-7): 841–50.

Daunay, M.-C. and Janick, J. (2007). History and Iconography of Eggplant, *Chronica Horticulturae* **47**(3): 16–22.

Daunay, M. and Hazra, P. (2012). Eggplant, *in* K. Peter and P. Hazra (eds), *Handbook of vegetables*, Studium Press, Houston, Texas, pp. 257–322.

Davidar, P., Snow, A. A., Rajkumar, M., Pasquet, R., Daunay, M.-C. and Mutegi, E. (2015). The potential for crop to wild hybridization in eggplant (*Solanum melongena*; Solanaceae) in southern India, *American Journal of Botany* **102**: 129–139.

del Rio, A. H., Bamberg, J. B. and Huaman, Z. (1997a). Assessing changes in the genetic diversity of potato gene banks. 1. Effects of seed increase, *TAG Theoretical and Applied Genetics* **95**(1-2): 191–198.

del Rio, A. H., Bamberg, J. B., Huaman, Z., Salas, A. and Vega, S. E. (1997b). Assessing changes in the genetic diversity of potato gene banks. 2. *In situ* vs *ex situ*, *TAG Theoretical and Applied Genetics* **95**(1-2): 199–204.

Delgado, C. L. (2003). Rising Consumption of Meat and Milk in Developing Countries Has Created a New Food Revolution, *The Journal of Nutrition* **133**(11): 3907S–3910.

Dempewolf, H., Eastwood, R. J., Guarino, L., Khoury, C. K., Müller, J. V. and Toll, J. (2013). Adapting agriculture to climate change: A global initiative to collect, conserve, and use crop wild relatives, *Agroecology and Sustainable Food Systems* **38**: 369–377.

Díez, M. and Nuez, F. (2008). Tomato, *in* J. Prohens and F. Nuez (eds), *Vegetables II SE - 7*, Vol. 2 of *Handbook of Plant Breeding*, Springer New York, pp. 249–323.

Doebley, J. F., Gaut, B. S. and Smith, B. D. (2006). The molecular genetics of crop domestication, *Cell* **127**: 1309–1321.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D. and Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, *Ecography* **36**(1): 27–46.

Dujardin, M. and Hanna, W. H. (1989). Crossability of Pearl Millet with Wild Pennisetum Species, *Crop Science* **29**(1): 77–80.

Eckert, C. G., Samis, K. E. and Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond, *Molecular ecology* **17**(5): 1170–88.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Townsend Peterson, A., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S. and Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data, *Ecography* **29**(2): 129–151.

Elith, J. and Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time, *Annual Review of Ecology, Evolution, and Systematics* **40**(1): 677–697.

Ellstrand, N. C., Prentice, H. C. and Hancock, J. F. (1999). Gene flow and introgression from domesticated plants into their wild relatives, *Annual Review of Ecology, Evolution, and Systematics* **30**: 539–563.

Esquinas-Alcázar, J. (2005). Protecting crop genetic diversity for food security: political, ethical and technical challenges, *Nature Reviews Genetics* **6**(12): 946–953.

ESRI (2011). ArcGIS Desktop: Release 10.1. Redlands, CA. Environmental Systems Research Institute.
URL: *http://www.esri.com/*

Estrada, N. (1991). Utilization of *Solanum brevidens* to transfer PLRV resistance into the cultivated potato, *Solanum tuberosum*, *in* J. Hawkes, R. Lester, M. Nee and N. Estrada (eds), *Solanaceae III: Taxonomy, Chemistry and Evolution*, Royal Botanical Gardens, Kew, London.

Estrada, R. N. (1980). Frost resistant potato hybrids via *Solanum acaule*, Bitt. Diploid - Tetraploid crosses, *American Potato Journal* **57**(12): 609–619.

European Space Agency (2009). GlobCover 2.2: Global land cover map.
URL: *http://due.esrin.esa.int/page_globcover.php*

Fajardo, D. and Spooner, D. M. (2011). Phylogenetic relationships of *Solanum* series Conicibaccata and related species in *Solanum* section Petota inferred from five conserved ortholog sequences, *Systematic Botany* **36**(1): 163–170.

FAO (2009). International treaty on plant genetic resources for food and agriculture.

FAO (2010). The second report on the state of the world's plant genetic resources for food and agriculture, *Technical report*, Rome, Italy.

FAO, IFAD and WFP (2015). The state of food insecurity in the world 2015. Meeting the 2015 international hunger targets: taking stock of uneven progress, *Technical report*, FAO, Rome.

FAO, WFP and IFAD (2012). *The State of Food Insecurity in the World. Economic growth is necessary but not sufficient to accelerate reduction of hunger and malnutrition*, FAO, Rome.

FAOSTAT (2015). FAOSTAT.
URL: *www.faostat.fao.org*

Farooq, S., Asghar, M., Iqbal, N., Askari, E., Arif, M. and Shah, T. M. (1995). Production of salt tolerant wheat germplasm through crossing cultivated wheat with *Aegilops cylindrica*-II. Field evaluation of salt tolerant germplasm, *Cereal Research Communications* **23**: 275–282.

Fielder, H., Brotherton, P., Hosking, J., Hopkins, J. J., Ford-Lloyd, B. and Maxted, N. (2015). Enhancing the conservation of crop wild relatives in England, *Plos One* **10**(6): e0130804.

Fielding, A. H. and Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models, *Environmental Conservation* **24**(01): 38–49.

Flor, A., Bettencourt, E., Arriegas, P. I. and Dias, S. R. (2006). European crop wild relative conservation criteria - indicators for the CWR species' list prioritization, *in* B. Ford-Lloyd, S. Dias and E. Bettencourt (eds), *Genetic Erosion and Pollution Assessment Methodologies. Proceedings of PGR Forum Workshop 5, Terceira Island, Autonomous Region of the Azores, Portugal, 8-11 September 2004.*, Published on behalf of the European Crop Wild Relative Diversity Assessment and Conservation Forum, Rome, Italy, pp. 83–88.

Foley, J. a., Defries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. a., Kucharik, C. J., Monfreda, C., Patz, J. a., Prentice, I. C., Ramankutty, N. and Snyder, P. K. (2005). Global consequences of land use., *Science* **309**(5734): 570–4.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D. and Zaks, D. P. M. (2011). Solutions for a cultivated planet, *Nature* **478**(7369): 337–342.

Food and Agriculture Organization of the United Nations (FAO). (2013). FAOSTAT.
URL: *www.faostat.fao.org*

Ford-Lloyd, B., Kell, S. P. and Maxted, N. (2008). Establishing conservation priorities for crop wild relatives, *in* N. Maxted, B. Ford-Lloyd, S. Kell, J. Iriondo, E. Dulloo and J. Turok (eds), *Crop wild relative conservation and use*, CABI, Wallingford, Oxfordshire, UK, chapter 6, pp. 110–119.

Ford-Lloyd, B. V., Schmidt, M., Armstrong, S. J., Barazani, O., Engels, J., Hadas, R., Hammer, K., Kell, S. P., Kang, D., Khoshbakht, K., Li, Y., Long, C., Lu, B.-R., Ma, K., Nguyen, V. T., Qiu, L., Ge, S., Wei, W., Zhang, Z. and Maxted, N. (2011). Crop wild relatives - Undervalued, underutilized and under threat?, *BioScience* **61**(7): 559–565.

Fowler, C., Moore, G. and Hawtin, G. (2003). *The international treaty on plant genetic resources for food and agriculture: a primer for the future harvest centres of the CGIAR*, International Plant Genetic Resources Institute, Rome, Italy.

Franklin, J. (2009). *Mapping species distributions: spatial inference and prediction*, Cambridge University Press, New York.

Franks, S. J., Weber, J. J. and Aitken, S. N. (2014). Evolutionary and plastic responses to climate change in terrestrial plant populations, *Evolutionary Applications* **7**(1): 123–139.

Frary, A., Doganlar, S. and Daunay, M. C. (2007). Eggplant, *in* C. Kole (ed.), *Vegetables SE - 9*, Vol. 5 of *Genome Mapping and Molecular Breeding in Plants*, Springer Berlin Heidelberg, Berlin, Germany, pp. 287–313.

Frost, K. E., Jansky, S. H. and Rouse, D. I. (2006). Transmission of *Verticillium* wilt resistance to tetraploid potato via unilateral sexual polyploidization, *Euphytica* **149**(3): 281–287.

Fuller, D. Q., Denham, T., Arroyo-Kalin, M., Lucas, L., Stevens, C. J., Qin, L., Allaby, R. G. and Purugganan, M. D. (2014). Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record, *Proceedings of the National Academy of Sciences of the United States of America* **111**(17): 6147–52.

Gaston, K. J. and Fuller, R. A. (2009). The sizes of species' geographic ranges, *Journal of Applied Ecology* **46**(1): 1–9.

Geldmann, J., Joppa, L. N. and Burgess, N. D. (2014). Mapping change in human pressure globally on land and within protected areas, *Conservation Biology* **28**(6): 1604–1616.

Gepts, P. (2006). Plant Genetic Resources Conservation and Utilization: The accomplishments and future of a societal insurance policy, *Crop Science* **46**(5): 2278–2292.

Girma, G., Spillane, C. and Gedil, M. (2016). DNA barcoding of the main cultivated yams and selected wild species in the genus *Dioscorea*, *Journal of Systematics and Evolution* **54**(3): 228–237.

Gisbert, C., Prohens, J., Raigón, M. D., Stommel, J. R. and Nuez, F. (2011). Eggplant relatives as sources of variation for developing new rootstocks: Effects of grafting on eggplant yield and fruit apparent quality and composition, *Scientia Horticulturae* **128**(1): 14–22.

Glover, J. D., Reganold, J. P., Bell, L. W., Borevitz, J., Brummer, E. C., Buckler, E. S., Cox, C. M., Cox, T. S., Crews, T. E., Culman, S. W., DeHaan, L. R., Eriksson, D., Gill, B. S., Holland, J., Hu, F., Hulke, B. S., Ibrahim, A. M. H., Jackson, W., Jones, S. S., Murray, S. C., Paterson, A. H., Ploschuk, E., Sacks, E. J., Snapp, S., Tao, D., Van Tassel, D. L., Wade, L. J., Wyse, D. L. and Xu, Y. (2010). Increased food and ecosystem security via perennial grains, *Science* **328**(5986): 1638–1639.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M. and Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people, *Science* **327**(5967): 812–818.

González, M., Galván, G., Siri, M. I., Borges, A. and Vilaró, F. (2013). Resistencia a la marchitez bacteriana de la papa en *Solanum commersonii*, *Agrociencia Uruguay* **7**: 45–54.

Goodwin, Z. A., Harris, D. J., Filer, D., Wood, J. R. and Scotland, R. W. (2015). Widespread mistaken identity in tropical plant collections, *Current Biology* **25**(22): R1066–R1067.

Gororo, N. N., Eagles, H. A., Eastwood, R. F., Nicolas, M. E. and Flood, R. G. (2002). Use of *Triticum tauschii* to improve yield of wheat in low-yielding environments, *Euphytica* **123**(2): 241–254.

Gramazio, P., Prohens, J., Plazas, M., Andújar, I., Herraiz, F. J., Castillo, E., Knapp, S., Meyer, R. S. and Vilanova, S. (2014). Location of chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant, *BMC Plant Biology* **14**(1): 350.

Guarino, L. and Lobell, D. B. (2011). A walk on the wild side, *Nature Climate Change* **1**(8): 374–375.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. a., Elith, J., Schwartz, M. W., Wintle, B. a., Broennimann, O., Austin, M., Ferrier, S., Kearney, M. R., Possingham, H. P. and Buckley, Y. M. (2013). Predicting species distributions for conservation decisions, *Ecology letters* **16**: 1424–1435.

Gur, A. and Zamir, D. (2004). Unused natural variation can lift yield barriers in plant breeding, *PLoS Biology* **2**(10): e245.

Hajjar, R. and Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years, *Euphytica* **156**(1-2): 1–13.

Hanneman, R. E. (1999). The reproductive biology of potato and its implications for breeding, *Potato Research* **42**(2): 283–312.

Harlan, J. R. and de Wet, J. M. J. (1971). Toward a rational classification of cultivated plants, *Taxon* **20**(4): 509–517.

Hart, R., Salick, J., Ranjitkar, S. and Xu, J. (2014). Herbarium specimens show contrasting phenological responses to Himalayan climate, *Proceedings of the National Academy of Sciences of the United States of America* **111**(29): 10615–10619.

Hawkes, J. G. (1979). Genetic poverty of the potato in Europe, *in* A. M. van Harten and A. C. Zeven (eds), *Broadening the genetic base of crops*, Centre for Agricultural Publishing and Documentation, Wageningen, pp. 19–27.

Hawkes, J. G., Maxted, N. and Ford-Lloyd, B. V. (2000). *The ex situ conservation of plant genetic resources*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Heisey, P. W. and Rubenstein, K. D. (2015). Using crop genetic resources to help agriculture adapt to climate change: Economics and policy, *Technical Report EIB-139*, U.S. Department of Agriculture, Economic Research Service.

Heywood, V., Casas, A., Ford-Lloyd, B., Kell, S. and Maxted, N. (2007). Conservation and sustainable use of crop wild relatives, *Agriculture, Ecosystems & Environment* **121**: 245–255.

Hijmans, R. J. (2014). *raster: raster: Geographic data analysis and modeling*.
URL: *http://cran.r-project.org/package=raster*

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

Hijmans, R. J., Garrett, K. a., Huamán, Z., Zhang, D. P., Schreuder, M. and Bonierbale, M. (2000). Assessing the Geographic Representativeness of Genebank Collections: the Case of Bolivian Wild Potatoes, *Conservation Biology* **14**(6): 1755–1765.

Hijmans, R. J., Gavrilenko, T., Stephenson, S., Bamberg, J., Salas, A. and Spooner, D. M. (2007). Geographical and environmental range expansion through polyploidy in wild potatoes (*Solanum* section Petota), *Global Ecology and Biogeography* **16**(4): 485–495.

Hijmans, R. J., Jacobs, M., Bamberg, J. B. and Spooner, D. M. (2003). Frost tolerance in wild potato species: Assessing the predictivity of taxonomic, geographic, and ecological factors, *Euphytica* **130**: 47–59.

Hijmans, R. J., Phillips, S., Leathwick, J. and Elith, J. (2014). *dismo: Species distribution modeling*.
URL: *http://cran.r-project.org/package=dismo*

Hijmans, R. J., Schreuder, M., De La Cruz, J. and Guarino, L. (1999). Using GIS to check co-ordinates of genebank accessions, *Genetic Resources and Crop Evolution* **46**: 291–296.

Hijmans, R. J. and Spooner, D. M. (2001). Geographic distribution of wild potato species, *American Journal of Botany* **88**(11): 2101–2112.

Hijmans, R. J., Spooner, D. M., Salas, A. R., Guarino, L. and de la Cruz, J. (2002). Atlas of Wild Potatoes, International Plant Genetic Resources Institute (IPGRI), Rome.

HLPE (2013). Investing in smallholder agriculture for food security. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security, *Technical Report June*, FAO, Rome.

Hodgkin, T. and Hajjar, R. (2008). Using crop wild relatives for crop improvement: trends and perspectives, *in* N. Maxted, B. Ford-Lloyd, S. Kell, J. Iriondo, E. Dulloo and J. Turok (eds), *Crop wild relatives conservation and use*, CABI Publishing, Wallingford, chapter 38, pp. 535–548.

Hoffmann, M. H., Glass, A. S., Tomiuk, J., Schmuths, H., Fritsch, R. M. and Bachmann, K. (2003). Analysis of molecular data of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS), *Molecular ecology* **12**(4): 1007–19.

Hoisington, D., Khairallah, M., Reeves, T., Ribaut, J. M., Skovmand, B., Taba, S. and Warburton, M. (1999). Plant genetic resources: what can they contribute toward increased crop productivity?, *Proceedings of the National Academy of Sciences of the United States of America* **96**(11): 5937–43.

Hudson, L. N., Newbold, T., Contu, S., Hill, S. L. L., Lysenko, I., Palma, D., Phillips, H. R. P., Senior, R. A., Bennett, D. J., Booth, H., Garon, M., Michelle, L., Correia, D. L. P., Day, J., Echeverr, S., Harrison, K., Ingram, D. J., Jung, M., Kemp, V., Kirkpatrick, L., Callum, D., Pan, Y., White, H. J., Aben, J., Abrahamczyk, S., Adum, G. B., Armbrecht, I., Azhar, B., Bates, A. J., Bayne, E. M., Beja, P., Banks, E., Barlow, J., Buczkowski, G., Buscardo, E., Cabra-garc, J., Cameron, S. A., Cancello, E. M., Carrijo, T. F., Carvalho, A. L., Castro-luna, A. A., Cerda, R., Cerezo, A., Chauvat, M., Darvill, B., Dauber, J., Dejean, A., Diek, T., Giordani, P., Giordano, S., Gottschalk, M. S., Goulson, D., Aaron, D., Grogan, J., Hanley, M. E., Hanson, T., Hashim, N. R., Joseph, E., Helden, A. J., Hern, L., Herzog, F., Higuera-diaz, D., Hilje, B., Horgan, F. G., Horv, R., Hylander, K., Isaacs-cubides, P., Ishitani, M., Jacobs, C. T., J, V., Jauker, B., Jonsell, M., Jung, T. S., Kapoor, V., Kati, V., Lantschner, V., F, V. L., Lebuhn, G., Jean-philippe, L., Louhaichi, M., Mallari, A., Marin-spiotta, E., Marshall, E. J. P., Mart, E., Muchane, M. N., Muchane, M., Naidoo, R., Nakamura, A., Naoe, S. and Fernando, A. B. (2014). The PREDICTS database: a global database of how local terrestrial biodiversity responds to human impacts, *Ecology and Evolution* **4**(24): 4701–4735.

IFAD and UNEP (2013). Smallholders, food security, and the environment, *Technical report*, Rome, Italy.

Integrated Taxonomic Information System (ITIS) (n.d.).
URL: *http://www.itis.gov/*

IPCC (2014). *Climate change 2014: Synthesis Report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*, IPCC, Geneva, Switzerland.

Iriondo, J. M., Maxted, N., Kell, S. P., Ford-lloyd, B. V., Lara-Romero, C., Labokas, J. and Magos Brehm, J. (2012). Quality standards for genetic reserve conservation of crop wild relatives, *in* N. Maxted, M. E. Dulloo, B. V. Ford-Lloyd, L. Frese, J. Iriondo and M. A. A. Pinheiro de Carvalho (eds), *Agrobiodiversity conservation: securing the diversity of crop wild relatives and landraces*, CABI, Wallingford, chapter 10, pp. 72–77.

IRRI (1990). Rice Germplasm: Collecting, Preservation, Use, *Proceedings of third international workshop*, Manila, Philippines, p. 185.

Irz, X., Lin, L., Thirtle, C. and Wiggins, S. (2001). Agricultural Productivity Growth and Poverty Alleviation, *Development Policy Review* **19**(4): 449–466.

Ishimaru, T., Hirabayashi, H., Ida, M., Takai, T., San-Oh, Y. a., Yoshinaga, S., Ando, I., Ogawa, T. and Kondo, M. (2010). A genetic resource for early-morning flowering trait of wild rice *Oryza officinalis* to mitigate high temperature-induced spikelet sterility at anthesis, *Annals of botany* **106**(3): 515–20.

IUCN (2012). IUCN Red List Categories and Criteria: Version 3.1., second edi edn, IUCN, Gland, Switzerland and Cambridge, UK, pp. iv + 32pp.

Jansky, S. H., Dempewolf, H., Camadro, E. L., Simon, R., Zimnoch-Guzowska, E., Bisognin, D. A. and Bonierbale, M. (2013). A case for crop wild relative preservation and use in potato, *Crop Science* **53**(3): 746–754.

Jansky, S. H., Hamernik, A. and Bethke, P. P. C. (2011). Germplasm release: Tetraploid clones with resistance to cold-induced sweetening, *American Journal of Potato Research* **88**(3): 218–225.

Jansky, S. H., Simon, R. and Spooner, D. M. (2006). A test of taxonomic predictivity: resistance to white mold in wild relatives of cultivated potato, *Crop Science* **46**: 2561–2570.

Jarl, C. I., Rietveld, E. M. and de Haas, J. M. (1999). Transfer of fungal tolerance through interspecific somatic hybridisation between *Solanum melongena* and *S. torvum*, *Plant Cell Reports* **18**(9): 791–796.

Jarret, R. (2008). DNA Barcoding in a crop genebank: The *Capsicum annuum* species complex, *The Open Biology Journal* **1**: 35–42.

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

Jarvis, A., Ramírez-Villegas, J., Vanessa, B., Campo, H. and Navarro-Racines, C. (2012). Is Cassava the Answer to African Climate Change Adaptation?, *Tropical Plant Biology* **5**: 9–29.

Jarvis, A., Williams, K., Williams, D., Guarino, L., Caballero, P. J. and Mottram, G. (2005). Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay, *Genetic Resources and Crop Evolution* **52**(6): 671–682.

Jo, K.-R., Arens, M., Kim, O.-Y., Jongsma, M. A., Visser, R. G. F., Jacobsen, E. and Vossen, J. H. (2011). Mapping of the *S. demissum* late blight resistance gene *R8* to a new locus on chromosome IX, *Theoretical and Applied Genetics* **123**(8): 1331–1340.

Johnson, C. J. and Gillingham, M. P. (2005). An evaluation of mapped species distribution models used for conservation planning, *Environmental Conservation* **32**(02): 117.

Johnson, J. A., Runge, C. F., Senauer, B., Foley, J. and Polasky, S. (2014). Global agriculture and carbon trade-offs, *Proceedings of the National Academy of Sciences of the United States of America* **2014**: 1–6.

Jones, J. D. G., Witek, K., Verweij, W., Jupe, F., Cooke, D., Dorling, S., Tomlinson, L., Smoker, M., Perkins, S. and Foster, S. (2014). Elevating crop disease resistance with cloned genes, *Philosophical Transactions of the Royal Society B* **369**: 20130087.

Kannenberg, L. W. and Falk, D. E. (1995). Models for activation of plant genetic resources for crop breeding programs, *Canadian Journal of Plant Science* **75**(1): 45–53.

Kantar, M. B., Sosa, C. C., Khoury, C. K., Castañeda-Álvarez, N. P., Achicanoy, H. a., Bernau, V., Kane, N. C., Marek, L., Seiler, G. and Rieseberg, L. H. (2015). Ecogeography and utility to plant breeding of the crop wild relatives of sunflower (*Helianthus annuus* L.), *Frontiers in Plant Science* **6**(841).

Kardos, M., Husby, A., McFarlane, S. E., Qvarnström, A. and Ellegren, H. (2015). Whole genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations, *Molecular ecology resources* .
URL: *http://dx.doi.org/10.1111/1755-0998.12498*

Kell, S. P., Maxted, N. and Bilz, M. (2012). European crop wild relative threat assessment: knowledge gained and lessons learnt, *in* N. Maxted, M. E. Dulloo, B. V. Ford-Lloyd, L. Frese, J. Iriondo and M. A. A. Pinheiro de Carvalho (eds), *Agrobiodiversity conservation: securing the diversity of crop wild relatives and landraces*, CABI, Wallingford, chapter 28, pp. 218–242.

Khan, M. and Isshiki, S. (2009). Functional male-sterility expressed in eggplant (*Solanum melongena* L.) containing the cytoplasm of *S. kurzii* Brace & Prain, *Journal of Horticultural Science and Biotechnology* **84**: 92–96.

Khoury, C. K., Greene, S., Wiersema, J., Maxted, N., Jarvis, A. and Struik, P. C. (2013). An Inventory of Crop Wild Relatives of the United States, *Crop Science* **53**(4): 1496.

Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., Rieseberg, L. H. and Struik, P. C. (2014). Increasing homogeneity in global food supplies and the implications for food security, *Proceedings of the National Academy of Sciences of the United States of America* **111**(11): 4001–4006.

Khoury, C. K., Heider, B., Castañeda-Álvarez, N. P., Achicanoy, H. a., Sosa, C. C., Miller, R. E., Scotland, R. W., Wood, J. R. I., Rossel, G., Eserman, L. a., Jarret, R. L., Yencho, G. C., Bernau, V., Juarez, H., Sotelo, S., de Haan, S. and Struik, P. C. (2015a). Distributions, *ex situ* conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., I. series Batatas], *Frontiers in Plant Science* **6**: 1–14.

Khoury, C. K., Castañeda-Álvarez, N. P., Achicanoy, H. A., Sosa, C. C., Bernau, V., Kassa, M. T., Norton, S. L., Maesen, L. J. G. V. D., Upadhyaya, H. D., Ramírez-Villegas, J., Jarvis, A. and Struik, P. C. (2015b). Crop wild relatives of pigeonpea [*Cajanus cajan* (L .) Millsp .]: Distributions, *ex situ* conservation status, and potential genetic resources for abiotic stress tolerance, *Biological Conservation* **184**: 259–270.

Khoury, C. K., Achicanoy, H. A., Bjorkman, A. D., Navarro-Racines, C., Guarino, L., Flores-Palacios, X., Engels, J. M. M., Wiersema, J. H., Dempewolf, H., Ramirez-Villegas, J., Castañeda-Álvarez, N. P., Fowler, C., Jarvis, A., Rieseberg, L. H. and Struik, P. C. (2015). Estimation of countries' interdependence in plant genetic resources provisioning national food supplies and production systems.
URL: *http://www.planttreaty.org/content/research-paper-8*

Knapp, S. and Vorontsova, M. S. (2013). From introduced American weed to Cape Verde Islands endemic: the case of *Solanum rigidum* Lam. (Solanaceae, *Solanum* subgenus *Leptostemonum*), *PhytoKeys* **46**(25): 35–46.

Knapp, S., Vorontsova, M. S. and Prohens, J. (2013). Wild relatives of the eggplant (*Solanum melongena* L.: Solanaceae): new understanding of species names in a complex group, *PloS one* **8**(2): e57039.

Kraak, A. (1992). Industrial applications of potato starch products, *Industrial Crops and Products* **1**(2-4): 107–112.

Kumchai, J., Wei, Y.-C., Lee, C.-Y., Chen, F.-C. and Chin, S.-W. (2013). Production of interspecific hybrids between commercial cultivars of the eggplant (*Solanum melongena* L.) and its wild relative *S. torvum*, *Genetics and Molecular Research* **12**(1): 755–64.

Laferriere, L. T., Helgeson, J. P. and Allen, C. (1999). Fertile *Solanum tuberosum* + *S. commersonii* somatic hybrids as sources of resistance to bacterial wilt caused by *Ralstonia solanacearum*, *TAG Theoretical and Applied Genetics* **98**(8): 1272–1278.

Lam, H.-M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.-L., Li, M.-W., He, W., Qin, N., Wang, B., Li, J., Jian, M., Wang, J., Shao, G., Wang, J., Sun, S. S.-M. and Zhang, G. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection, *Nature genetics* **42**(12): 1053–1059.

Lester, R. and Hasan, S. (1991). Origin and domestication of the brinjal eggplant, *Solanum melongena*, from *Solanum incanum*, in Africa and Asia, *in* J. G. Hawkes, R. N. Lester, M. Nee and N. Estrada (eds), *Solanaceae III: Taxonomy, Chemistry and Evolution*, Royal Botanic Gardens, Kew, London, pp. 369–387.

Lin, Z., Eaves, D. J., Sanchez-Moran, E., Franklin, F. C. H. and Franklin-Tong, V. E. (2015). The *Papaver rhoeas* S determinants confer self-incompatibility to *Arabidopsis thaliana* in planta, *Science* **350**(6261): 684–687.

Linares, O. F. (2002). African rice (*Oryza glaberrima*): history and future potential, *Proceedings of the National Academy of Sciences of the United States of America* **99**(25): 16360–16365.

Lindqvist-Kreuze, H., Carbajulca, D., Gonzáles-Escobedo, J., Pérez, W. and Bonierbale, M. (2010). Comparison of transcript profiles in late blight-challenged *Solanum cajamarquense* and B3C1 potato clones, *Molecular Plant Pathology* **11**(4): 513–530.

Liu, C., Berry, P. M., Dawson, T. P. and Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions, *Ecography* **28**(3): 385–393.

Liu, J., Zheng, Z., Zhou, X., Feng, C. and Zhuang, Y. (2015). Improving the resistance of eggplant (*Solanum melongena*) to *Verticillium* wilt using wild species *Solanum linnaeanum*, *Euphytica* **201**(3): 463–469.

Liu, Z. and Halterman, D. (2009). Different Genetic Mechanisms Control Foliar and Tuber Resistance to *Phytophthora infestans* in Wild Potato *Solanum verrucosum*, *American Journal of Potato Research* **86**(6): 476–480.

Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P. and Naylor, R. L. (2008). Prioritizing climate change adaptation needs for food security in 2030, *Science* **319**(5863): 607–10.

López Noriega, I., Halewood, M., Galluzzi, G., Vernooy, R., Bertacchini, E., Gauchan, D. and Welch, E. (2013). How policies affect the use of plant genetic resources: the experience of the CGIAR, *Resources* **2**: 231–269.

Luthra, S. K., Gopal, J., Kumar, D., Singh, B. P. and Pandey, S. K. (2009). *Solanum* wild and cultivated species as source of resistance to cold induced sweetening, *Potato Journal* **36**(3/4): 115–120.

Mace, G. M., Collar, N. J., Gaston, K. J., Hilton-Taylor, C., Akçakaya, H. R., Leader-Williams, N., Milner-Gulland, E. J. and Stuart, S. N. (2008). Quantification of extinction risk: IUCN's system for classifying threatened species, *Conservation Biology* **22**(6): 1424–1442.

Magos Brehm, J., Maxted, N., Ford-Lloyd, B. V. and Martins-Loução, M. A. (2008). National inventories of crop wild relatives and wild harvested plants: case-study for Portugal, *Genetic Resources and Crop Evolution* **55**(6): 779–796.

Mariac, C., Luong, V., Kapran, I., Mamadou, A., Sagnard, F., Deu, M., Chantereau, J., Gerard, B., Ndjeunga, J., Bezançon, G., Pham, J.-L. and Vigouroux, Y. (2006). Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum* [L.] R. Br.) in Niger assessed by microsatellite markers, *TAG. Theoretical and applied genetics* **114**(1): 49–58.

Mattheij, W. M., Eijlander, R., de Koning, J. R. and Louwes, K. M. (1992). Interspecific hybridization between the cultivated potato *Solanum tuberosum* subspecies *tuberosum* L. and the wild species *S. circaeifolium* subsp. *circaeifolium* Bitter exhibiting resistance to *Phytophthora infestans* (Mont.) de Bary and *Globodera pallida*, *Theoretical and applied genetics* **83**(4): 459–66.

Maxted, N., Dulloo, E., Ford-Lloyd, B. V., Iriondo, J. M. and Jarvis, A. (2008). Gap analysis: a tool for complementary genetic conservation assessment, *Diversity and Distributions* **14**(6): 1018–1030.

Maxted, N., Ford-Lloyd, B. V., Jury, S., Kell, S. and Scholten, M. (2006). Towards a definition of a crop wild relative, *Biodiversity and Conservation* **15**(8): 2673–2685.

Maxted, N., Hargreaves, S., Kell, S. P., Amri, A., Street, K., Shehadeh, A., Piggin, J. and Konopka, J. (2012). Temperate forage and pulse legume genetic gap analysis, *Bocconea* pp. 115–146.

Maxted, N. and Kell, S. (2009). *Establishment of a global network for the in situ conservation of crop wild relatives: status and needs*, FAO Commission on Genetic Resources for Food and Agriculture, Italy, Rome.

Maxted, N., Kell, S. and Brehm, J. M. (2014). Crop Wild Relatives and Climate Change, *in* M. Jackson, B. Ford-Lloyd and M. Parry (eds), *Plant Genetic Resources and Climate Change*, CABI International, Wallingford UK, chapter 7, pp. 114–136.

McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., Burke, J. M., Charest, D., Cloutier, S., Cole, G., Dempewolf, H., Dingkuhn, M., Feuillet, C., Gepts, P., Grattapaglia, D., Guarino, L., Jackson, S., Knapp, S., Langridge, P., Lawton-Rauh, A., Lijua, Q., Lusty, C., Michael, T., Myles, S., Naito, K., Nelson, R. L., Pontarollo, R., Richards, C. M., Rieseberg, L., Ross-Ibarra, J., Rounsley, S., Hamilton, R. S., Schurr, U., Stein, N., Tomooka, N., van der Knaap, E., van Tassel, D., Toll, J., Valls, J., Varshney, R. K., Ward, J., Waugh, R., Wenzl, P. and Zamir, D. (2013). Feeding the future, *Nature* **499**: 23–24.

McCouch, S. R., McNally, K. L., Wang, W. and Hamilton, R. S. (2012). Genomics of gene banks: A case study in rice, *American Journal of Botany* **99**(2): 407–423.

McCouch, S. R., Sweeney, M., Li, J., Jiang, H., Thomson, M., Septiningsih, E., Edwards, J., Moncada, P., Xiao, J., Garris, A., Tai, T., Martinez, C., Tohme, J., Sugiono, M., McClung, A., Yuan, L. P. and Ahn, S. N. (2007). Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*, *Euphytica* **154**(3): 317–339.

Meilleur, B. a. and Hodgkin, T. (2004). *In situ* conservation of crop wild relatives: status and trends, *Biodiversity and Conservation* **13**(4): 663–684.

Mendoza, H. A. and Haynes, F. L. (1974). Genetic relationship among potato cultivars grown in the United States, *HortScience* **9**: 328–330.

Mennella, G., Rotino, G. L., Fibiani, M., D'Alessandro, A., Francese, G., Toppino, L., Cavallanti, F., Acciarri, N. and Lo Scalzo, R. (2010). Characterization of health-related compounds in eggplant (*Solanum melongena* L.) lines derived from introgression of allied species, *Journal of Agricultural and Food Chemistry* **58**(13): 7597–7603.

Merow, C., Smith, M. J. and Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter, *Ecography* **36**(10): 1058–1069.

Meyer, C., Kreft, H., Guralnick, R. and Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions, *Nature Communications* **6**: 8221.

Meyer, R. S., Karol, K. G., Little, D. P., Nee, M. H. and Litt, A. (2012a). Phylogeographic relationships among Asian eggplants and new perspectives on eggplant domestication, *Molecular phylogenetics and evolution* **63**(3): 685–701.

Meyer, R. S., DuVal, A. E. and Jensen, H. R. (2012b). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops, *The New Phytologist* **196**(1): 29–48.

Miller, J. and Seiler, G. (2003). Registration of Five Oilseed Maintainer (HA 429-HA 433) Sunflower Germplasm Lines, *Crop Science* **43**(6): 2313.

Mondoni, A., Rossi, G., Orsenigo, S. and Probert, R. J. (2012). Climate warming could shift the timing of seed germination in alpine plants, *Annals of Botany* **110**(1): 155–164.

Moss, J. P., Singh, A. K., Reddy, L. J., Nigam, S. N., Subrahmanyam, P., McDonald, D. and Reddy, A. G. S. (1997). Registration of ICGV 87165 Peanut Germplasm Line with Multiple Resistance, *Crop Science* **37**: 1028.

Msangi, S. and Rosegrant, M. W. (2011). Feeding the future's changing diets. Implications for agriculture markets, nutrition and policy, *2020 Conference: leveraging agriculture for improving nutrition and health*, number February, New Delhi, India.

Mutegi, E., Snow, A. A., Rajkumar, M., Pasquet, R., Ponniah, H., Daunay, M.-C. and Davidar, P. (2015). Genetic diversity and population structure of wild/weedy eggplant (*Solanum insanum*, Solanaceae) in southern India: Implications for conservation, *American Journal of Botany* **102**(1): 140–148.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. and Kent, J. (2000). Biodiversity hotspots for conservation priorities, *Nature* **403**(6772): 853–8.

Myers, S. S., Zanobetti, A., Kloog, I., Huybers, P., Leakey, A. D. B., Bloom, A. J., Carlisle, E., Dietterich, L. H., Fitzgerald, G., Hasegawa, T., Holbrook, N. M., Nelson, R. L., Ottman, M. J., Raboy, V., Sakai, H., Sartor, K. A., Schwartz, J., Seneweera, S., Tausz, M. and Usui, Y. (2014). Increasing CO2 threatens human nutrition, *Nature* **510**(7503): 139–42.

Nabholz, B., Sarah, G., Sabot, F., Ruiz, M., Adam, H., Nidelet, S., Ghesquière, A., Santoni, S., David, J. and Glémin, S. (2014). Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*), *Molecular Ecology* **23**(9): 2210–27.

Naess, S. K., Bradeen, J. M., Wielgus, S. M., Haberlach, G. T., McGrath, J. M. and Helgeson, J. P. (2000). Resistance to late blight in *Solanum bulbocastanum* is mapped to chromosome 8, *TAG Theoretical and Applied Genetics* **101**(5-6): 697–704.

Narancio, R., Zorrilla, P., Robello, C., Gonzalez, M., Vilaró, F., Pritsch, C. and Rizza, M. (2013). Insights on gene expression response of a characterized resistant genotype of *Solanum commersonii* Dun. against *Ralstonia solanacearum*, *European Journal of Plant Pathology* **136**(4): 823–835.

Nass, L. L. and Paterniani, E. (2000). Pre-breeding: a link between genetic resources and maize breeding, *Scientia Agricola* **57**(3): 581–587.

Nix, H. A. (1986). A biogeographic analysis of Australian elapid snakes, *in* R. Longmore (ed.), *Atlas of Elapid Snakes of Australia*, australian edn, Australian Government Publishing Service, Canberra, pp. 4–15.

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

Ochoa, C. (1999). Las papas de Sudamerica: Perú, Allen Press, Lawrence, USA, p. 1036.

Ochoa, C. (2003). Las papas del Perú: base de datos 1947-1997, Universidad Nacional Agraria La Molina (UNALM), Agencia Suiza para el Desarrollo y la Cooperación (COSUDE), Centro Internacional de la Papa (CIP), Lima, Perú.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P. and Kassem, K. R. (2001). Terrestrial ecoregions of the world: a new map of life on Earth, *Bioscience* **51**(11): 933–938.

Olson, M. B., Morris, K. S. and Méndez, V. E. (2012). Cultivation of maize landraces by small-scale shade coffee farmers in western El Salvador, *Agricultural Systems* **111**: 63–74.

Ovchinnikova, A., Krylova, E., Gavrilenko, T., Smekalova, T., Zhuk, M., Knapp, S. and Spooner, D. (2011). Taxonomy of cultivated potatoes (*Solanum* section Petota: Solanaceae), *Botanical Journal of the Linnean Society* **165**: 107–155.

Parr, C. S., Wilson, N., Leary, P., Schulz, K., Lans, K., Walley, L., Hammock, J., Goddard, A., Rice, J., Studer, M., Holmes, J. and Corrigan Jr., R. J. (2014). The Encyclopedia of Life v2: Providing global access to knowledge about life on Earth, *Biodiversity Data Journal* **2**: e1079.

Parra-Quijano, M., Iriondo, J. M. and Torres, E. (2011). Improving representativeness of genebank collections through species distribution models, gap analysis and ecogeographical maps, *Biodiversity and Conservation* **21**(1): 79–96.

Paulsen, G. M. (2000). International contributions to the improvement and marketing of Kansas wheat.

Pavek, J. J. and Corsini, D. L. (2001). Utilization of potato genetic resources in variety development, *American Journal of Potato Research* **78**(6): 433–441.

Pearson, R. G., Raxworthy, C. J., Nakamura, M. and Townsend Peterson, a. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar, *Journal of Biogeography* **34**(1): 102–117.

Pebesma, E. and Bivand, R. (2005). Classes and methods for spatial data in R. R News 5 (2). **URL:** *http://cran.r-project.org/doc/Rnews/*

Pelletier, Y. (2007). Level and genetic variability of resistance to the colorado potato beetle (*Leptinotarsa decemlineata* (Say)) in wild *Solanum* species, *American Journal of Potato Research* **84**(2): 143–148.

Perring, M., Cullen, B., Johnson, I. and Hovenden, M. (2010). Modelled effects of rising CO2 concentration and climate change on native perennial grass and sown grass-legume pastures, *Climate Research* **42**(1): 65–78.

Peterson, A., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M. and Araújo, M. B. (2011). *Ecological niches and geographic distributions*, Princeton University Press, New Jersey, USA.

Peterson, A. T., Navarro-Siguenza, A. G., Martinez-Meyer, E., Cuervo-Robayo, A. P., Berlanga, H. and Soberón, J. (2015). Twentieth century turnover of Mexican endemic avifaunas: Landscape change versus climate drivers, *Science Advances* **1**(4): e1400071.

Peterson, A. T., Ortega-Huerta, M. A., Bartley, J., Sánchez-Cordero, V., Soberón, J., Buddemeier, R. H. and Stockwell, D. R. B. (2002). Future projections for Mexican faunas under global climate change scenarios, *Nature* **416**(6881): 626–9.

Pfeiffer, W. H. and McClafferty, B. (2007). HarvestPlus: Breeding Crops for Better Nutrition, *Crop Science* **47**.

Phillips, S. J., Anderson, R. P. and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions, *Ecological Modelling* **190**(3-4): 231–259.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. and Ferrer, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data, *Ecological Applications* **19**(1): 181–197.

Pimentel, D., Wilson, C., McCullum, C., Huang, R., Dwen, P., Flack, J., Tran, Q., Saltman, T. and Cliff, B. (1997). Economic and environmental benefits of biodiversity, *BioScience* **47**(11): 747–757.

Pistorius, R. (1997). *Scientists, plants and politics - A history of the plant genetic resources movement*, International Plant Genetic Resources Institute (IPGRI), Rome, Italy.

Plazas, M., López-Gresa, M. P., Vilanova, S., Torres, C., Hurtado, M., Gramazio, P., Andújar, I., Herráiz, F. J., Bellés, J. M. and Prohens, J. (2013). Diversity and relationships in key traits for functional and apparent quality in a collection of eggplant: Fruit phenolics content, antioxidant activity, polyphenol oxidase activity, and browning, *Journal of Agricultural and Food Chemistry* **61**(37): 8871–8879.

Plazas, M., Vilanova, S., Gramazio, A., Rodriguez-Burruezo, A., Fita, F., Herraiz, F., Ranil, R. and Fonseka, R. (2016). Interspecific hybridization between eggplant and wild relatives from different genepools, *Journal of the American Society for Horticultural Science* **141**: 34–44.

Porter, J., Xie, L., Challinor, A., Cochrane, K., Howden, S., Iqbal, M., Lobell, D. and Travasso, M. (2014). Food security and food production systems, *in* C. Field, V. Barros, D. Dokken, K. Mach, M. Mastrandrea, T. Bilir, M. Chatterjee, K. Ebi, Y. Estrada, R. Genova, B. Girma, E. Kissel, A. Levy, S. MacCracken, P. Mastrandrea and L. White (eds), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, chapter 7, pp. 485–533.

Pringle, R. M., Goheen, J. R., Palmer, T. M., Charles, G. K., Defranco, E., Ford, A. T. and Tarnita, C. E. (2014). Low functional redundancy among mammalian browsers in regulating an encroaching shrub (*Solanum campylacanthum*) in African savannah, *Proceedings of the Royal Society B: Biological Sciences* p. 20140390.

Prischmann, D. a., Dashiell, K. E., Schneider, D. J. and Eubanks, M. W. (2009). Evaluating *Tripsacum* -introgressed maize germplasm after infestation with western corn rootworms (Coleoptera: Chrysomelidae), *Journal of Applied Entomology* **133**(1): 10–20.

Prohens, J., Anderson, G. J., Herraiz, F. J., Bernardello, G., Santos-Guerra, A., Crawford, D. and Nuez, F. (2007). Genetic diversity and conservation of two endangered eggplant relatives (*Solanum vespertilio* Aiton and *Solanum lidii* Sunding) endemic to the Canary Islands, *Genetic Resources and Crop Evolution* **54**(3): 451–464.

Prohens, J., Whitaker, B. D., Plazas, M., Vilanova, S., Hurtado, M., Blasco, M., Gramazio, P. and Stommel, J. R. (2013). Genetic diversity in morphological characters and phenolic acids content resulting from an interspecific cross between eggplant, Solanum melongena, and its wild ancestor (*S. incanum*), *Annals of Applied Biology* **162**(2): 242–257.

PWC (2013). Crop wild relatives: A valuable resource for crop development, *Technical Report July*.

R Core Team (2014). R: A language and environment for statistical computing.
URL: *http://www.r-project.org/*

Rahman, M., Rashid, M., Hossain, M., Salam, M. and Masum, A. (2002). Grafting compatibility of cultivated eggplant varieties with wild *Solanum* species, *Pakistan Journal of Biological Science* **5**(7): 755–757.

Raigón, M. D., Prohens, J., Muñoz-Falcón, J. E. and Nuez, F. (2008). Comparison of eggplant landraces and commercial varieties for fruit content of phenolics, minerals, dry matter and protein, *Journal of Food Composition and Analysis* **21**(5): 370–376.

Ramirez, J., Jarvis, A., Bergh, I. V. D., Staver, C. and Turner, D. (2011). Changing Climates: Effects on Growing Conditions for Banana and Plantain (*Musa* spp.) and Possible Responses, *in* S. S. Yadav, R. Redden, J. L. Hatfield, H. Lotze-Campen and A. J. W. Hall (eds), *Crop Adaptation to Climate Change*, Wiley-Blackwell, chapter Chapter 20, pp. 426–438.

Ramirez-Villegas, J. and Jarvis, A. (2010). Downscaling global circulation outputs: The delta method. Decision and Policy Analysis Working Paper No. 1.

Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D. G. and Guarino, L. (2010). A gap analysis methodology for collecting crop genepools: a case study with *Phaseolus* beans, *PloS one* **5**(10): e13497.

Rao, G. R. and Kumar, A. (1980). Some observations on interspecific hybrids of *Solanum melongena* L, *Proceedings of the Indian Academy of Sciences (Plant Science)* **89**: 117–121.

Ray, D. K., Mueller, N. D., West, P. C. and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050, *PLoS ONE* **8**(6): e66428.

Reif, J. C., Zhang, P., Dreisigacker, S., Warburton, M. L., Van Ginkel, M., Hoisington, D., Bohn, M. and Melchinger, a. E. (2005). Wheat genetic diversity trends during domestication and breeding, *Theoretical and Applied Genetics* **110**(5): 859–864.

Richardson, D. M. and Rejmánek, M. (2011). Trees and shrubs as invasive alien species - a global review, *Diversity and Distributions* **17**(5): 788–809.

Rick, C. M. and Chetelat, R. T. (1995). Utilization of related wild species for tomato improvement, *in* R. Fernández-Muñoz, J. Cuartero and M. Gómez-Guillamón (eds), *I International Symposium on Solanaceae for Fresh Market*, ISHS, Malaga, pp. 21–38.

Rios, N. E. and Bart, H. L. (2010). GEOLocate.
URL: *http://www.museum.tulane.edu/geolocate/*

Robbirt, K. M., Roberts, D. L., Hutchings, M. J. and Davy, A. J. (2014). Potential disruption of pollination in a sexually deceptive orchid by climatic change, *Current Biology* **24**(23): 2845–2849.

Rodrigues, A. S. L., Pilgrim, J. D., Lamoreux, J. F., Hoffmann, M. and Brooks, T. M. (2006). The value of the IUCN Red List for conservation, *Trends in Ecology & Evolution* **21**(2): 71–76.

Rodríguez Delgado, O., García Gallo, A. and Cruz Trujillo, G. (2011). *Solanum lidii*.
URL: *www.iucnredlist.org*

Rodriguez, F., Wu, F., Ané, C., Tanksley, S. and Spooner, D. M. (2009). Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history?, *BMC evolutionary biology* **9**(1): 191.

Rosegrant, M. W. and Cai, X. (2001). Water scarcity and food security: alternative futures for the 21st century, *Water Science and Technology* **43**(4): 61–70.

Rosegrant, M. W., Cai, X. and Cline, S. A. (2002). Global water outlook to 2025: Averting an Impending Crisis, *Technical report*, International Food Policy Research Institute (IFPRI), Washington.

Ross, H. (1979). Wild species and primitive cultivars as ancestors of potato varieties, *in* A. Zeven and A. van Harten (eds), *Broadening the genetic base of crops*, Centre for Agricultural Publishing and Documentation, Wageningen, pp. 237–245.

Ross, H. (1986). Potato breeding - Problems and perspectives, *in* P. Parey (ed.), *Fortschritte der Pflanzenzüchtung 13*, Berlin and Hamburg.

Rotino, G. L., Sala, T. and Toppino, L. (2014). Eggplant, *in* A. Pratap and J. Kumar (eds), *Alien Gene Transfer in Crop Plants, Volume 2 SE - 16*, Springer New York, pp. 381–409.

Rudorf, W. (1958). The significance of wild species for potato breeding, *European Potato Journal* **1**: 10–20.

Ruiz-Vera, U. M., Siebers, M., Gray, S. B., Drag, D. W., Rosenthal, D. M., Kimball, B. a., Ort, D. R. and Bernacchi, C. J. (2013). Global warming can negate the expected CO2 stimulation in photosynthesis and productivity for soybean grown in the Midwestern United States, *Plant physiology* **162**(1): 410–23.

Russell, J., van Zonneveld, M., Dawson, I. K., Booth, A., Waugh, R. and Steffenson, B. (2014). Genetic diversity and ecological niche modelling of wild barley: refugia, large-scale post-LGM range expansion and limited mid-future climate threats?, *PLoS ONE* **9**(2): e86021.

Santini, M., Camadro, E. L., Marcellan, O. N. and Erazzu, L. U. (2000). Agronomic characterization of diploid hybrid families derived from crosses between haploids of the common potato and three wild Argentinian tuber-bearing species, *American Journal of Potato Research* **77**: 211–218.

Sarikamis, G., Marquez, J., Maccormack, R., Bennett, R. N., Roberts, J. and Mithen, R. (2006). High glucosinolate broccoli: a delivery system for sulforaphane, *Molecular Breeding* **18**: 219–228.

Särkinen, T., Bohs, L., Olmstead, R. G. and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree., *BMC evolutionary biology* **13**(1): 214.

Särkinen, T., Iganci, J. R. V., Linares-Palomino, R., Simon, M. F. and Prado, D. E. (2011). Forgotten forests - issues and prospects in biome mapping using Seasonally Dry Tropical Forests as a case study, *BMC ecology* **11**(27).

Sboner, A., Mu, X. J., Greenbaum, D., Auerbach, R. K. and Gerstein, M. B. (2011). The real cost of sequencing: higher than you think!, *Genome Biology* **12**(8): 125.

Schafleitner, R., Ramirez, J., Jarvis, A., Evers, D., Gutierrez, R. and Scurrah, M. (2011). Adaptation of the Potato Crop to Changing Climates, *in* S. S. Yadav, R. Redden, J. L. Hatfield, H. Lotze-Campen and A. J. W. Hall (eds), *Crop Adaptation to Climate Change*, Wiley-Blackwell, chapter Chapter 11, pp. 287–297.

Schalk, J. M., Stoner, A. K. and RE Webb, H. F. W. (1975). Resistance in eggplant, *Solanum melongena* L. and nontuber-bearing Solanum species to carmine spider mite, *Journal of the American Society for Horticultural Science* **100**(5): 479–481.

Schlenker, W. and Lobell, D. B. (2010). Robust negative impacts of climate change on African agriculture, *Environmental Research Letters* **5**(1): 014010.

Secretariat of the CBD (2010). Decision X/2. The Strategic Plan for Biodiversity 2011-2020 and the Aichi Biodiversity Targets.
URL: *http://www.biodiv.be/convention/strategic-plan-2011-2020/*

Secretariat of the CBD (2011). *Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the Convention on Biological Diversity: text and annex*, number October, Montreal.

Seddon, A. W. R., Macias-Fauria, M., Long, P. R., Benz, D. and Willis, K. J. (2016). Sensitivity of global terrestrial ecosystems to climate variability, *Nature* pp. 1–15.

Settele, J., Scholes, R., Betts, R. A., Bunn, S., Leadley, P., Nepstad, D., Overpeck, J. T. and Taboada, M. A. (2014). Terrestrial and Inland Water Systems, *in* C. Field, V. Barros, D. Dokken, K. Mach, M. Mastrandrea, T. Bilir, M. Chatterjee, K. Ebi, Y. Estrada, R. Genova, B. Girma, E. Kissel, A. Levy, S. MacCracken, P. Mastrandrea and L. White (eds), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, chapter 4, pp. 271–359.

Sharma, S., Upadhyaya, H. D., Varshney, R. K. and Gowda, C. L. L. (2013). Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes, *Frontiers in plant science* **4**: 1–14.

Simon, R., Fuentes, A. F. and Spooner, D. M. (2011). Biogeographic Implications of the Striking Discovery of a 4,000 Kilometer Disjunct Population of the Wild Potato *Solanum morelliforme* in South America, *Systematic Botany* **36**(4): 1062–1067.

Simon, R., Xie, C. H., Clausen, A., Jansky, S. H., Halterman, D., Conner, T., Knapp, S., Brundage, J., Symon, D. and Spooner, D. (2010). Wild and cultivated potato (*Solanum* sect. Petota) escaped and persistent outside of its natural range, *Invasive Plant Science and Management* **3**(3): 286–293.

Singh, M. and Kumar, R. (2007). Eggplant (melongena L.), *in* R. J. Singh (ed.), *Genetic Resources, Chromosome Engineering, and Crop Improvement. Vegetable Crops, Volume 3*, CRC Press-Taylor & Francis Group, Boca Raton, FL, chapter 10, pp. 473–495.

Singh, S., Mackill, D. J. and Ismail, A. M. (2009). Responses of SUB1 rice introgression lines to submergence in the field: Yield and grain quality, *Field Crops Research* **113**(1): 12–23.
URL: *http://www.sciencedirect.com/science/article/pii/S037842900900094X*

Slavin, J. and Lloyd, B. (2012). Health benefits of fruits and vegetables, *Advances in Nutrition* **3**(4): 506–516.

Smith, N. J. H. and Schultes, R. E. (1990). Deforestation and Shrinking Crop Gene-pools in Amazonia, *Environmental Conservation* **17**(3): 227–234.

Smyda, P., Jakuczun, H., Debski, K., Sliwka, J., Thieme, R., Nachtigall, M., Wasilewicz-Flis, I. and Zimnoch-Guzowska, E. (2013). Development of somatic hybrids *Solanum* x *michoacanum* Bitter. (Rydb.) (+) *S. tuberosum* L. and autofused 4x *S.* x *michoacanum* plants as potential sources of late blight resistance for potato breeding, *Plant cell reports* **32**(8): 1231–41.

Spooner, D. M. (2009). DNA barcoding will frequently fail in complicated groups: An example in wild potatoes, *American Journal of Botany* **96**(6): 1177–1189.

Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H. and Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes, *The Botanical Review* **80**(4): 283–383.

Spooner, D. M. and Hijmans, R. J. (2001). Potato systematics and germplasm collecting, 1989 - 2000, *American Journal of Potato Research* **78**(4): 237–268.

Stephenson, N. L. (1998). Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales, *Journal of Biogeography* **25**: 855–870.

Stern, S., Agra, M. D. F. and Bohs, L. (2011). Molecular delimitation of clades within New World species of the "spiny solanums" (*Solanum* subg. *Leptostemonum*), *Taxon* **60**(5): 1429–1441.

Stevenson, W., Loria, R., Franc, G. D. and Weingartner, D. P. (2001). *Compendium of potato diseases*, 2nd editio edn, American Phytopathological Society Press, St Paul, USA.

Suárez, S., Chaves, E., Clausen, A. and Franco, J. (2009). *Solanum* tuber-bearing species resistance behavior against *Nacobbus aberrans*, *Journal of Nematology* **41**(1): 5–10.

Subbarao, G. V., Nakahara, K., Hurtado, M. P., Ono, H., Moreta, D. E., Salcedo, a. F., Yoshihashi, a. T., Ishikawa, T., Ishitani, M., Ohnishi-Kameyama, M., Yoshida, M., Rondon, M., Rao, I. M., Lascano, C. E., Berry, W. L. and Ito, O. (2009). Evidence for biological nitrification inhibition in *Brachiaria* pastures, *Proceedings of the National Academy of Sciences of the United States of America* **106**(41): 17302–17307.

Suneson, C. A. (1967). Registration of Sierra Oats1 (Reg. No. 213), *Crop Science* **7**(2): 168.

Sunseri, F., Sciancalepore, A., Martelli, G., Acciarri, N., Rotino, G., Valentino, D. and Tamietti, G. (2003). Development of a RAPD-AFLP map of eggplant and improvement of tolerance to *Verticillium* wilt, *Acta Horticulturae* **625**: 197–198.

Syfert, M. M., Castañeda-Álvarez, N. P., Khoury, C. K., Särkinen, T., Sosa, C. C., Achicanoy, H. A., Bernau, V., Prohens, J., Daunay, M.-C. and Knapp, S. (2016). Crop wild relatives of the brinjal eggplant (*Solanum melongena*: Solanaceae): poorly represented in genebanks and many species at risk of extinction, *American journal of botany* .

Syfert, M. M., Smith, M. J. and Coomes, D. A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models, *PLoS ONE* **8**(2): e55158.

Syngenta (2013). Our industry 2013, *Technical report*, Basel, Switzerland.

Tanksley, S. D. and McCouch, S. R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild, *Science* **277**: 1063–1066.

Taub, D. R., Miller, B. and Allen, H. (2008). Effects of elevated CO2 on the protein concentration of food crops: a meta-analysis, *Global Change Biology* **14**(3): 565–575.

The World Bank (2014a). Agriculture, value added (% of GDP): National accounts data.
URL: *http://data.worldbank.org/indicator/NV.AGR.TOTL.ZS*

The World Bank (2014b). Employment in agriculture (% of total employment).
URL: *http://data.worldbank.org/indicator/SL.AGR.EMPL.ZS*

The International Plant Names Index (2012).
URL: *http://www.ipni.org/*

Thieme, R., Rakosy-Tican, E., Gavrilenko, T., Antonova, O., Schubert, J., Nachtigall, M., Heimbach, U. and Thieme, T. (2008). Novel somatic hybrids (*Solanum tuberosum* L.+ *Solanum tarnii*) and their fertile BC1 progenies express extreme resistance to potato virus Y and late blight., *TAG. Theoretical and applied genetics* **116**(5): 691–700.

Thiers, B. (2010). Index Herbariorum: A global directory of public herbaria and associated staff.
URL: *http://sweetgum.nybg.org/ih/*

Thirtle, C., Lin, L. and Piesse, J. (2003). The Impact of Research-Led Agricultural Productivity Growth on Poverty Reduction in Africa, Asia and Latin America, *World Development* **31**(12): 1959–1975.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., De Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Peterson, A. T., Phillips, O. L. and Williams, S. E. (2004). Extinction risk from climate change, *Nature* **427**: 145–148.

Thomas, E., van Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G. and van Etten, J. (2012). Present spatial diversity patterns of *Theobroma cacao* L. in the neotropic reflect genetic differentiation in Pleistocene refugia followed by human-influenced dispersal, *PLoS ONE* **7**(10): e47676.

Thomson, A. M., Calvin, K. V., Smith, S. J., Kyle, G. P., Volke, A., Patel, P., Delgado-Arias, S., Bond-Lamberty, B., Wise, M. A., Clarke, L. E. and Edmonds, J. A. (2011). RCP4.5: A pathway for stabilization of radiative forcing by 2100, *Climatic Change* **109**: 77–94.

Thuiller, W., Lavergne, S., Roquet, C., Boulangeat, I., Lafourcade, B. and Araujo, M. B. (2011). Consequences of climate change on the tree of life in Europe, *Nature* **470**(7335): 531–4.

Tilman, D., Balzer, C., Hill, J. and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture, *Proceedings of the National Academy of Sciences of the United States of America* **108**(50): 20260–20264.

Tilman, D. and Clark, M. (2014). Global diets link environmental sustainability and human health, *Nature* **515**(7528): 518–522.

Tittensor, D. P., Walpole, M., Hill, S. L. L., Boyce, D. G., Britten, G. L., Burgess, N. D., Butchart, S. H. M., Leadley, P. W., Regan, E. C., Alkemade, R., Baumung, R., Bellard, C., Bouwman, L., Bowles-newark, N. J., Chenery, A. M. and Cheung, W. W. L. (2014). Biodiversity Targets, *Science* **346**(6206): 241–245.

Traka, M. H., Saha, S., Huseby, S., Kopriva, S., Walley, P. G., Barker, G. C., Moore, J., Mero, G., van den Bosch, F., Constant, H., Kelly, L., Schepers, H., Boddupalli, S. and Mithen, R. F. (2013). Genetic regulation of glucoraphanin accumulation in Beneforté broccoli, *The New phytologist* **198**(4): 1085–95.

Tucci, M., Carputo, D., Bile, G. and Frusciante, F. (1996). Male fertility and freezing tolerance of hybrids involving *Solanum tuberosum* haploids and diploid *Solanum* species, *Potato Research* **39**: 345–353.

Tyack, N. and Dempewolf, H. (2015). The economics of crop wild relatives under climate change, *in* R. Redden, S. S. Yadav, N. Maxted, M. E. Dulloo, L. Guarino and P. Smith (eds), *Crop Wild Relatives and Climate Change*, Wiley Blackwell, New Jersey, chapter 16, pp. 281–291.

Uhrig, H. and Wenzel, G. (1981). *Solanum gourlayi* Hawkes as a source of resistance against the white potato cyst nematode *Globodera pallida* Stone, *Zeitschrift für Pflanzenzüchtung* **86**: 148–157.

Uitdewillgen, J., Wolters, A., D'Hoop, B., Borm, T., Visser, R. and van Eck, H. (2013). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato, *PloS one* **8**(5): 1–14.

United Nations (2013). World Population Prospects: The 2012 Revision, Key Findings and Advance Tables.

United Nations (2015). World Population Prospects 2015.
URL: *http://www.un.org/esa/population/publications/WPP2004/2004Highlights_finalrevised.pdf*

Ureta, C., Martínez-Meyer, E., Perales, H. R. and Álvarez-Buylla, E. R. (2012). Projecting the effects of climate change on the distribution of maize races and their wild relatives in Mexico, *Global Change Biology* **18**(3): 1073–1082.

USDA ARS National Genetic Resources Program (n.d.). Germplasm Resources Information Network - (GRIN).
URL: *https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomysearch.aspx*

Usher, M. (2000). The nativeness and non-nativeness of species, *Watsonia* **23**(2): 323–326.

Valkoun, J. J. (2001). Wheat pre-breeding using wild progenitors, *Euphytica* **119**(1-2): 17–23.

van de Wouw, M., Kik, C., van Hintum, T., van Treuren, R. and Visser, B. (2009). Genetic erosion in crops: concept, research results and challenges, *Plant Genetic Resources* **8**(01): 1–15.

van der Vossen, E., Sikkema, A., Hekkert, B. t. L., Gros, J., Stevens, P., Muskens, M., Wouters, D., Pereira, A., Stiekema, W. and Allefs, S. (2003). An ancient R gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato., *The Plant journal: for cell and molecular biology* **36**(6): 867–82.

Van Vuuren, D., Bouwman, A. and Beusen, A. (2010). Phosphorus demand for the 1970-2100 period: A scenario analysis of resource depletion, *Global Environmental Change* **20**(3): 428–439.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J. and Rose, S. K. (2011). The representative concentration pathways: an overview, *Climatic Change* **109**(1-2): 5–31.

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L. and Storlie, C. (2014). *SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises*.
URL: *http://cran.r-project.org/package=SDMTools*

VanDerWal, J., Shoo, L. P., Graham, C. and Williams, S. E. (2009). Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know?, *Ecological Modelling* **220**(4): 589–594.

Vasconcelos, A. C. F., Bonatti, M., Schlindwein, S. L., D'Agostini, L. R., Homem, L. R. and Nelson, R. (2013). Landraces as an adaptation strategy to climate change for smallholders in Santa Catarina, Southern Brazil, *Land Use Policy* **34**: 250–254.

Vavilov, N. I. (1926). Centers of origin of cultivated plants, *Bulletin of applied botany and plant breeding* **16**(2).

Vavilov, N. I. (1951). The origin, variation, immunity and breeding of cultivated plants, *Chronica Botanica* **13**: 1–366.

Vavilov, N. I. (1992). *Origin and geography of cultivated crops*, Cambridge University Press, London, UK.

Vice-ministry for the environment biodiversity and climate change (VMABCC) and Bioversity International (2009). Red book of crop wild relatives in Bolivia, Plural Editores, La Paz, p. 340.

Villamon, F., Spooner, D., Orrillo, M., Mihovilovich, E., Pérez, W. and Bonierbale, M. (2005). Late blight resistance linkages in a novel cross of the wild potato species *Solanum paucissectum* (series Piurana), *Theoretical and Applied Genetics* **111**: 1201–1214.

Vimal, R., Rodrigues, A. S. L., Mathevet, R. and Thompson, J. D. (2010). The sensitivity of gap analysis to conservation targets, *Biodiversity and Conservation* **20**(3): 531–543.

Vincent, H., Castañeda-Álvarez, N. P. and Maxted, N. (n.d.). Global priorities for *in situ* conservation of wild plant genetic resources: towards the establishment of a worldwide network of crop wild relative reserves.

Vincent, H., von Bothmer, R., Knüpffer, H., Amri, A., Konopka, J. and Maxted, N. (2012). Genetic gap analysis of wild *Hordeum* taxa, *Plant Genetic Resources* **10**(03): 242–253.

Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N. P., Guarino, L., Eastwood, R., León, B. and Maxted, N. (2013). A prioritized crop wild relative inventory to help underpin global food security, *Biological Conservation* **167**: 265–275.

Vorontsova, M. S., Kirka, P. and Muthoka, P. (2010). Overlooked diversity in African *Solanum* (Solanaceae): new and endangered *Solanum agnewiorum* from Kenya, *Phytotaxa* **10**: 31–37.

Vorontsova, M. S. and Knapp, S. (in press). A revision of the spiny solanums, *Solanum* subgenus *Leptostemonum* (Solanaceae) in Africa and Madagascar, *Systematic Botany Monographs* .

Vorontsova, M. S. and Mbago, F. M. (2010). Species from Tanzanian coastal forests may already be extinct, *Journal of East African Natural History* **99**(2): 227–234.

Vorontsova, M. S., Stern, S., Bohs, L. and Knapp, S. (2013). African spiny *Solanum* (subgenus *Leptostemonum*, Solanaceae): a thorny phylogenetic tangle, *Botanical Journal of the Linnean Society* **173**: 176–193.

Wale, S. J., Platt, H. W. and Cattlin, N. D. (2008). Diseases, Pests and Disorders of Potatoes: A Color Handbook, Elsevier, Florida, p. 176.

Wang, F., Li, F., Wang, J., Zhou, Y. and Sun, H. (2011). Genetic Diversity of the Selected 64 Potato Germplasms Revealed by AFLP Markers, *Molecular Plant Breeding* **2**: 22–29.

Wang, J.-X., Gao, T.-G. and Knapp, S. (2008). Ancient Chinese literature reveals pathways of eggplant domestication, *Annals of botany* **102**(6): 891–7.

Warren, R., VanDerWal, J., Price, J., Welbergen, J. a., Atkinson, I., Ramirez-Villegas, J., Osborn, T. J., Jarvis, a., Shoo, L. P., Williams, S. E. and Lowe, J. (2013). Quantifying the benefit of early climate change mitigation in avoiding biodiversity loss, *Nature Climate Change* **3**(5): 1–5.

Watanabe, K., Orrillo, M., Vega, S., Masuelli, R. and Ishiki, K. (1994). Potato germ plasm enhancement with disomic tetraploid *Solanum acaule*. II. Assessment of breeding value of tetraploid F1 hybrids between tetrasomic tetraploid *S. tuberosum* and *S. acaule*, *Theoretical and Applied Genetics* **88**: 135–140.

Watanabe, K., Orrillo, M., Vega, S., Valkonen, J., Pehu, E., Hurtado, A. and Tanksley, S. (1995). Overcoming crossing barriers between non tuber-bearing and tuber-bearing *Solanum* species: towards potato germplasm enhancement with a broad spectrum of solanaceous genetic resources, *Genome* **38**(1): 27–35.

Watson, J. E. M., Iwamura, T. and Butt, N. (2013). Mapping vulnerability and conservation adaptation strategies under climate change, *Nature Climate Change* **3**(9): 1–6.

Webb, D. (1985). What are the criteria for presuming native status?, *Watsonia* **15**: 231–236.

Weese, T. L. and Bohs, L. (2007). A three-gene phylogeny of the genus *Solanum* (Solanaceae), *Systematic Botany* **32**(2): 445–463.

Weese, T. L. and Bohs, L. (2010). Eggplant origins: Out of Africa, into the Orient, *Taxon* **59**(1): 49–56.

White, F. (1983). *The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa*, Vol. 20, United Nations, Switzerland.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*, Springer New York.

Wiersema, J. H., León, B. and Garvey, E. J. (2012). Identifying wild relatives of subtropical and temperate fruit and nut crops, *in* M. Aradhya and D. Kluepfel (eds), *ISHS Acta Horticulturae 948: I International Symposium on Wild Relatives of Subtropical and Temperate Fruit and Nut Crops*, International Society for Horticultural Science, The Hague, The Netherlands, pp. 285–288.

Wilkes, G. (2007). Urgent notice to all maize researchers: disappearance and extinction of the last wild *Teosinte* population is more than half completed. A modest proposal for *Teosinte* evolution and conservation *in situ*: The Balsas, Guerrero, Mexico, *Maydica* **52**: 49–58.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A. and The Nceas Predicting Species Distributions Working Group (2008). Effects of sample size on the performance of species distribution models, *Diversity and Distributions* **14**: 763–773.

World Database on Protected Areas Consortium (2010). World Database on Protected Areas. **URL:** *http://www.protectedplanet.net/*

Wusirika, R., Li, K., Phillips, R. L. and Bennetzen, J. L. (2011). Zea, *in* C. Kole (ed.), *Wild Crop Relatives: Genomic and Breeding Resources, Cereals*, Springer-Verlag, Berlin, Heidelberg, chapter 11.

Xiao, J., Grandillo, S., Ahn, S. N., McCouch, S. R. and Tanksley, S. D. (1996). Genes from wild rice improve yield, *Nature* **384**: 356–358.

Yadav, K. (2012). Genetic Diversity of Pigeonpea (*Cajanus cajan* (L.) Millsp.) Cultivars and Its Wild Relatives Using Randomly Amplified Polymorphic DNA (RAPD) Markers, *American Journal of Plant Sciences* **03**(03): 322–330.

Yadav, S. S., Kumar, J., Turner, N. C., Berger, J., Redden, R., McNeil, D., Materne, M., Knights, E. J. and Bahl, P. N. (2004). Breeding for improved productivity, multiple resistance and wide adaptation in chickpea (*Cicer arietinum* L.), *Plant Genetic Resources: characterization and utilization* **2**(3): 181–187.

Zhang, X., Zhou, S., Fu, Y., Su, Z., Wang, X. and Sun, C. (2006). Identification of a drought tolerant introgression line derived from Dongxiang common wild rice (*O. rufipogon* Griff.), *Plant molecular biology* **62**(1-2): 247–59.

Zhu, Y., Chen, H., Fan, J., Wang, Y., Li, Y., Chen, J., Fan, J., Yang, S., Hu, L., Leung, H., Mew, T. W., Teng, P. S., Wang, Z. and Mundt, C. C. (2000). Genetic diversity and disease control in rice, *Nature* **406**(6797): 718–22.

Zimmerer, K. S. (2013). The compatibility of agricultural intensification in a global hotspot of smallholder agrobiodiversity (Bolivia), *Proceedings of the National Academy of Sciences of the United States of America* **110**(8): 2769–74.