

**Genetic diversity of cacao (*Theobroma cacao* L.) populations and
agronomic traits for breeding**

Dissertation

To attain the Degree Doctor of Philosophy (Ph.D.)
of the Faculty of Forest Sciences and Forest Ecology

Georg-August-Universität Göttingen

Submitted by

Helmuth Edisson Nieves-Orduña

born in Bogotá, D.C., Colombia

Göttingen, December 2023

1st Referee	Prof. Dr. Oliver Gailing
2nd Referee	Prof. Dr. Ralph Mitlöhner
Date of oral examination	28.11.2023

*“El equilibrio empieza a dar su luz, cuando arriba en la
montaña baila un guerrero noches enteras”.*

-El equilibrio parte I, Jaguares

ACKNOWLEDGEMENTS

I am happy to thank those who made this project possible. I am indebted to my supervisors; I extend my heartfelt thanks to Prof. Dr. Oliver Gailing for his continuous support from the beginning of this project until its completion. Thank you for the valuable suggestions on how to conduct my research, the opportunity to work in the lab, and the guidance during data analysis and manuscript writing. Your time and support were essential to the success of this research. I thank Prof. Dr. Konstantin V. Krutovsky for his suggestions and discussions during our seminars and for his data analysis and manuscript improvement advice. I thank Prof. Dr. Ralph Mitlöhner for participating in the thesis committee meetings and for suggesting to me how to organize my work. I am grateful to PD Dr. Markus Müller for assisting me with the data analysis and manuscript refining. Furthermore, many thanks to the lab members Alexandra Dolynska, Lioba Degenhardt, Gerold Dinkel, and Christine Radler. In particular, I appreciate Dr. Oliver Caré and Dr. Thi Ha Giang Vu for the MassARRAY training. Also, I thank the Dr. Carina CM Moura, Blessing Durodola, Sayed Jalal Moosavi and Andrii Tarieiev for sharing their expertise and fruitful conversations.

I would like to thank Dr. Andrew Daymond and Stella Polle of The International Cocoa Quarantine Centre, at the University of Reading (ICQC,R) for providing the cacao leaf samples. Likewise, I would also like to thank the team of the International Cacao *ex situ* collection managed by the Tropical Agricultural Research and Higher Education Center (CATIE) for collecting and sending the cacao samples from Turrialba, Costa Rica, to Göttingen. Additionally, I extend my appreciation to Mr. Allan Mata for making this collaboration possible especially during the challenging times of the global pandemic.

I am grateful to all the speakers and participants of the 2017 Cacao for Peace Research Symposium in Barranquilla, Colombia, where I engaged in inspiring conversations with the international cacao research community. I would like to express my gratitude to Prof. Wilbert

Phillips Mora of CATIE for sharing his research results and advising me to apply to the DAAD (Deutscher Akademischer Austauschdienst).

I sincerely thank the DAAD for their financial support, which made this research possible. I am also grateful to the GFA (Graduate School Forest and Agricultural Sciences) for additional funding. Furthermore, I would like to acknowledge Göttingen University, Faculty of Forest Sciences and Forest Ecology, for providing me with the necessary laboratory and office space for this research.

A la comunidad Latina de Gotinga, muchas gracias por compartir buenos momentos. Cristian, Victor Guerra (qepd) y Yazan; Natalia, Laura y Daniel (the running club); Martin (el juglar); Santiago y David (the swimming club), Andres T. y Jhon Z. ¡Con mucho gusto! nos volveremos a encontrar.

Sin la ayuda incondicional de mi familia este proyecto no hubiese sido posible. Gracias infinitas por todo el amor, por compartir a pesar de la distancia durante este periodo lleno de tantos retos. A mis hermanos Valentino, Alejandro y Néstor. A mi padre Jesús Nieves y a mi madre Cecilia Orduña por estar ahí, tan firmes como el sol, cuando más los necesite. Gracias a Dios.

Gracias mil...

Göttingen, September 2023.

PREFACE

This dissertation consists of a general introduction to cacao and the use of genetic resources for crop improvement, then three core research papers are presented, two published and the third submitted to an international peer-reviewed journal. It concludes with a general discussion that presents the main findings, conclusions, and research gaps related to conserving and using cacao genetic resources for crop breeding.

The first scientific paper, Geographic distribution, conservation, and genomic resources of cacao, *Theobroma cacao* L., analyses the geographic distribution of cacao and pinpoints new areas for germplasm collection. It discusses the conservation of cacao genetic resources, the genetic basis of agronomic traits, and genomic resources for cacao improvement.

The second paper, Geographic patterns of genetic variation among cacao (*Theobroma cacao* L.) populations based on chloroplast markers, focuses on the phylogeographic structure of natural cacao populations; chloroplast DNA (cpDNA) polymorphisms were used to show how chloroplast genetic diversity is geographically distributed in Amazonia, and how it is represented in common cacao genotypes of farms and breeding populations.

The third paper, Characterization of cacao (*Theobroma cacao* L.) accessions based on published SNP markers linked to agronomic traits reveals signals of selection, provides information on population structure of wild and cultivated cacao, identifies agronomic traits under selection, and uncovers valuable cacao genotypes not observed in cultivated and breeding populations.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
PREFACE	iii
General introduction	1
1 The chocolate tree and its genetic diversity	1
2 Domestication and breeding of cacao	2
3 References	6
Geographic distribution, conservation, and genomic resources of cacao	
<i>Theobroma cacao</i> L.	11
1 Introduction	15
2 Geographic distribution of natural populations	17
3 <i>Ex situ</i> and <i>in situ</i> conservation of cacao	22
4 Genetic basis of agronomic traits	29
5 Genomic resources available for cacao breeding	49
6 Conclusions and outlook	51
7 References	57
Geographic patterns of genetic variation among cacao (<i>Theobroma cacao</i> L.)	
populations based on chloroplast SSR markers	80
1 Introduction	84
2 Materials and Methods	87
3 Results	91
4 Discussion.....	98
5 Conclusions	104
6 Outlook	104
7 References	107
8 Supplementary Materials.....	115
Genotyping of cacao (<i>Theobroma cacao</i> L.) germplasms resources with SNP markers	
linked to agronomic traits reveals signs of selection	116
1 Introduction	120
2 Materials and Methods	122
3 Results	126
4 Discussion.....	134
5 Conclusions and research questions	144
6 References	148

GENERAL DISCUSSION	155
CONCLUSIONS	161
OUTLOOK	162
SUMMARY	164
ZUSAMMENFASSUNG	167

LIST OF ABBREVIATIONS

AMOVA	Analysis of Molecular Variance
BPR	Black Pod Rot
CATIE	International Center for Tropical Agriculture
Cd	Cadmium
CSE	Caffeoyl-Shikimate Esterase
FAs	Fatty Acids
FPR	Frosty Pod Rot
GS	Genomic Selection
GWAS	Genome-Wide Association Studies
HMA	Heavy Metal ATPase
ICGD	International Cocoa Germplasm Database
ICGT	International Cocoa Genebank Trinidad
LRR	Leucine-Rich Repeat
NB	Nucleotide-Binding
NRAMP5	Natural Resistance-Associated Macrophage Proteins
PCoA	Principal Coordinates Analysis
PR	Pathogenesis-Related
QTL	Quantitative Trait Loci
SC	Self-Compatibility
SI	Self-Incompatibility
SNP	Single-Nucleotide Polymorphism
SSR	Simple Sequence Repeats;
cpSSR	chloroplast SSR
WBD	Witches' Broom disease

GENERAL INTRODUCTION

1 THE CHOCOLATE TREE AND ITS GENETIC DIVERSITY

The cacao tree (*Theobroma cacao* L.), a diploid species ($2n = 2x = 20$) with a small genome (430–445 Mbp), belongs to the Malvaceae family and is native to Amazonia (Argout et al., 2011; Motamayor et al., 2013). Cacao is a shade-tolerant and relatively small tree. Its small, bisexual flowers are pollinated by midges, and the cauliflorous fruits have a hard and rugose surface, containing 20-40 seeds surrounded by sweet pulp (Wood & Lass, 1985). Cacao is cultivated in the tropics mainly by smallholder farmers under agroforestry systems to produce seeds, the unique source of chocolate (Coe & Coe, 2013; Daymond & Bekele, 2022; ICCO, 2021). A higher demand from the global chocolate industry drives increments in cacao cultivation globally, top producers include west Africa (76%), Latin America (17%) and Asia (6%) (FAOSTAT, 2020; ICCO, 2021). The global cacao production in 2020/21 was 5175 thousand tons of seeds (ICCO, 2021), major per capita consumers of chocolate include Europe and the US (Lindt & Sprüngli, 2018).

Western Amazonia is the hot spot of cacao genetic diversity. Seven out of eleven genetic clusters identified with nuclear microsatellite markers are located in the Ecuadorian-Peruvian Amazon (Motamayor et al., 2008; Zhang et al., 2012). This area also holds the highest variation of chloroplast haplotypes when compared to eastern Amazonia (Nieves-Orduña et al., 2023). Future collection trips likely will reveal new genetic clusters, due to high biological diversity of western Amazonia (Bass et al., 2010; Steege et al., 2003; Tuomisto et al., 1995). Unfortunately, due to persistent deforestation in Amazonia, the cacao primary gene pool is losing habitat, which likely is reducing the population size of natural populations (Smith & Schultes, 1990; Soares-Filho et al., 2006; Steege et al., 2015). Although there have been collection trips for wild material in Amazonia since the 1930s (Zhang & Motilal, 2016), areas

not yet explored for cacao *ex situ* collections with high probability of capturing new genetic diversity include river systems at the upper border between Ecuador-Peru (Nieves-Orduña et al., 2023). In addition, this area is important for collecting wild material that likely coevolved resistance to the damaging Frosty Pod Rot disease (caused by *Moniliophthora roreri*) (Díaz-Valderrama et al., 2022).

2 DOMESTICATION AND BREEDING OF CACAO

Domestication of cacao occurred ~3600 years ago in South America (Cornejo et al., 2018), with evidence of cacao use in Ecuador dating ~5600 years ago thanks to the recovery of ancient cacao DNA at the archeological site Santa Ana – La Floridad of Ecuador (Zarrillo et al., 2018). The domestication of cacao, like other fruit tree species of Amazonia, likely targeted traits such as pulp flavor, which was consumed directly but also used for the preparation of fermented beverages (Clement et al., 2010; Venturieri, 2013). Traits such as reduced bitterness were selected by native Americans in ancient Criollo populations (Cornejo et al., 2018). In addition, cacao domestication led to the accumulation of deleterious mutations affecting fitness and yield in Criollo (Cornejo et al., 2018).

Criollo cacao was cultivated during colonial times, but modern cacao varieties are mainly hybrid varieties that preserve the traditional cacao flavor (Bartley, 2005; Cornejo et al., 2018). Populations genetic analysis based on 200 genomes revealed that cultivated cacao and breeding populations are mainly composed of hybrids with a high ancestry of Criollo and Amelonado clusters (Cornejo et al., 2018), confirming the underutilization of cacao genetic resources for developing breeding populations and cultivars (Zhang & Motilal, 2016). Since the beginning of the cacao trade in the seventeenth century to present day, the fine flavor attributed to Criollo ancestry has had a higher price at the international markets (Dand, 2011; ICCO, 2022).

The narrow genetic diversity of cultivated Criollo led to a series of historical reductions in cacao production likely due to the lack of disease resistance traits (Díaz-Valderrama et al., 2020; Zhang & Motilal, 2016). Examples include the cacao blast in 1727 in Trinidad, where mostly Criollo was cultivated, which forced the first introduction of wild cacao (called Forastero) to broaden the genetic base of cacao farms (Zhang et al., 2011). The first cacao breeding program took place in Trinidad and targeted yield and high seed weight while preserving the fine chocolate flavor (Toxopeus, 1969). Three old diseases destroyed up to 38% of global crop production and have been the focus of cacao breeding: Black Pod-BP (caused by *Phytophthora* spp.), Witches Broom Disease-WBD (caused by *M. pernicioso*), Frosty Pod Rot-FPR (caused by *M. royeri*) (Gutiérrez et al., 2016; Marelli et al., 2019). Traditional cross breeding has successfully preserved chocolate attributes while combining disease resistance and yield traits sourced from selected wild cacao (Boza et al., 2014; Goenaga et al., 2009; Phillips-Mora et al., 2013).

Modern genomic resources in cacao breeding include the analysis of two reference genomes Criollo B97-61 and Matina1-6, and characterization of key genes related to fat biosynthesis, fine flavor, and fruit color (Argout et al., 2011; Motamayor et al., 2013). The Infinium cacao 6K/15K SNP array is available for molecular breeding (Livingstone et al., 2015; Livingstone et al., 2017), the EMBRAPA multispecies chip, which includes 3.4K of these cacao SNPs, has made genotyping more affordable at around USD\$20 per sample (Lopes et al., 2022). The cacao SNP array has been used in QTL mapping (Gutiérrez, Puig, et al., 2021; Royaert et al., 2016), GWAS studies (Romero-Navarro et al., 2017), genomic selection (McElroy et al., 2018) and estimation of genome-wide SNP homozygosity for developing inbred lines and exploiting heterosis (Lopes et al., 2022). In addition, cacao SNPs to facilitate the selection process are available for agronomic traits such as disease resistance to FPR, BP, and WBD (Gutiérrez, Puig, et al., 2021; Royaert et al., 2016), and yield (Bekele et al., 2022;

Fernandes et al., 2020). A set of 216 SNPs selected over the ten cacao chromosomes is available to classify and identify germplasm accessions (Gutiérrez, Martínez, et al., 2021).

Breeding in cacao still is a challenge but the primary gene pool holds novel alleles for present and future breeding activities (Zhang & Motilal, 2016). Traits of economic importance include disease resistance (especially to FPR), yield, reduced seed cadmium content, reduced tree height, sexual compatibility, and flavor (Bartley, 2005; Bekele & Phillips-Mora, 2019; Gutiérrez et al., 2016; Lanaud et al., 2017). Up to date the conservation of cacao genetic resources is organized in 40 *ex situ* collections (over 24k accessions are maintained), including two international collections (CacaoNet, 2012). In addition, challenges remain for the management of *ex situ* collections such as identification of duplicated and mislabeled accessions, genetic gaps, standardization of agronomic phenotyping, protocols for cryopreservation, and financial support to ensure long term conservation, where the chocolate industry can play an important role (CacaoNet, 2012; Nieves-Orduña et al., 2023).

Planting genetically improved varieties secures crop productivity and decreases pressure for new farmland, which reduces the risk of deforestation linked to cacao cultivation, and secures economic benefits for farmers and sustainability for the chocolate industry (Bekele & Phillips-Mora, 2019; Gutiérrez et al., 2016; Morrissey et al., 2019). Efficient conservation and use of underutilized cacao genetic resources are fundamental to address current and future breeding objectives.

This research aims to study the genetic diversity, conservation, and utilization of cacao genetic resources for crop improvement. Two types of genetic markers were used to study the genetic diversity of cacao populations (wild, cultivated, and breeding populations), cacao chloroplast microsatellite (simple sequence repeat - SSR) markers and published SNP (single nucleotide polymorphism) markers associated with economically important agronomic traits

such as disease resistance and yield. A comprehensive review analyzing the conservation and utilization of cacao genomic resources for cacao breeding is also presented.

The specific objectives of this dissertation are:

- to identify the potential geographic distribution of wild cacao populations in Amazonia and show patterns of geographic genetic variation,
- to assess the conservation status and available cacao genomic resources for cultivar development,
- to review the genetic basis of agronomic traits such as disease resistance, quality traits and yield, and available molecular markers for cultivar development,
- to establish the geographic distribution of chloroplast (cpDNA) variation in different populations representing natural cacao stands, cultivated cacao, and breeding populations,
- to describe the population structure and analyze genetic clusters in wild and managed cacao, and identify underutilized genetic resources for cacao breeding,
- to identify agronomic traits under selection that differentiate wild and managed cacao to assist breeding of superior cacao genotypes.

3 REFERENCES

- Argout, X., Salse, J., Aury, J.-M., Gaultinan, M., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S., Abrouk, M., Murat, F., Fouet, O., Poulain, J., Ruiz, M., Roguet, Y., Rodier-Goud, M., Barbosa-Neto, J. F., Sabot, F., . . . Lanaud, C. (2011). The genome of *Theobroma cacao*. *Nature Genetics*, 43(2), 101–108. <https://doi.org/10.1038/ng.736>
- Bartley, B. G. D. (2005). *The genetic diversity of cacao and its utilization*. CABI Pub.
- Bass, M. S., Finer, M., Jenkins, C. N., Kreft, H., Cisneros-Heredia, D. F., McCracken, S. F., Pitman, N. C. A., English, P. H., Swing, K., Villa, G., Di Fiore, A., Voigt, C. C., & Kunz, T. H. (2010). Global conservation significance of Ecuador's Yasuní National Park. *PloS One*, 5(1), e8767. <https://doi.org/10.1371/journal.pone.0008767>
- Bekele, F., Bidaisee, G., Allegre, M., Argout, X., Fouet, O., Boccara, M., Saravanakumar, D., Bekele, I., & Lanaud, C. (2022). Genome-wide association studies and genomic selection assays made in a large sample of cacao (*Theobroma cacao* L.) germplasm reveal significant marker-trait associations and good predictive value for improving yield potential. *PloS One*, 17(10), e0260907.
- Bekele, F., & Phillips-Mora, W. (2019). Cacao (*Theobroma cacao* L.) breeding. In J. M. Al-Khayri, S. M. Jain, & D. V. Johnson (Eds.), *Advances in Plant Breeding Strategies: Industrial and Food Crops* (Vol. 46, pp. 409–487). Springer International Publishing. https://doi.org/10.1007/978-3-030-23265-8_12
- Boza, E. J., Motamayor, J. C., Amores, F. M., Cedeño-Amador, S., Tondo, C. L., Livingstone, D. S., Schnell, R. J., & Gutiérrez, O. A. (2014). Genetic characterization of the cacao cultivar CCN 51: its impact and significance on global cacao improvement and production. *Journal of the American Society for Horticultural Science*, 139(2), 219–229. <https://doi.org/10.21273/JASHS.139.2.219>
- CacaoNet. (2012). *A global strategy for the conservation and use of cacao genetic resources, as the foundation for a sustainable cocoa economy*. Montpellier, France. Bioversity International.
- Clement, C., Cristo-Araújo, M. de, Coppens D'Eeckenbrugge, G., Alves Pereira, A., & Picanço-Rodrigues, D. (2010). Origin and domestication of native Amazonian crops. *Diversity*, 2(1), 72–106. <https://doi.org/10.3390/d2010072>
- Coe, S. D., & Coe, M. D. (2013). *The true history of chocolate* (Third edition). Thames & Hudson.
- Cornejo, O. E., Yee, M.-C., Dominguez, V., Andrews, M., Sockell, A., Strandberg, E., Livingstone, D., Stack, C., Romero, A., Umaharan, P., Royaert, S., Tawari, N. R., Ng, P., Gutierrez, O., Phillips, W., Mockaitis, K., Bustamante, C. D., & Motamayor, J. C. (2018). Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process. *Communications Biology*, 1, 167. <https://doi.org/10.1038/s42003-018-0168-6>
- Dand, R. (2011). *The international cocoa trade* (Third Edition). Elsevier. <https://doi.org/10.1016/C2013-0-16163-6>
- Daymond, A., & Bekele, F. (2022). Cacao. In P. M. Priyadarshan & S. M. Jain (Eds.), *Cash Crops: Genetic diversity, erosion, conservation and utilization* (pp. 23–53). Springer International Publishing. https://doi.org/10.1007/978-3-030-74926-2_2

- Díaz-Valderrama, J. R., Leiva-Espinoza, S. T., & Aime, M. C. (2020). The history of cacao and its diseases in the Americas. *Phytopathology*, *110*(10), 1604–1619. <https://doi.org/10.1094/PHYTO-05-20-0178-RVW>
- Díaz-Valderrama, J. R., Zambrano, R., Cedeño-Amador, S., Córdova-Bermejo, U., Casas, G. G., García-Zurita, N., Sánchez-Arévalo, J. A. J., Arévalo-Gardini, E., Dávila, D., Ruiz, J., Pinchi-Dávila, X., Quispe-Chacón, Z. R., Chia-Wong, J. A., Hurtado-Gonzales, O. P., Rodríguez-Callañaupa, C. A., Maldonado-Fuentes, C., Pérez-Callizaya, E., Leiva-Espinoza, S., Oliva-Cruz, M., . . . Aime, M. C. (2022). Diversity in the invasive cacao pathogen *Moniliophthora roreri* is shaped by agriculture. *Plant Pathology*, Article ppa.13603. Advance online publication. <https://doi.org/10.1111/ppa.13603>
- FAOSTAT. (2020). *Crops and livestock products: Cocoa, beans*. <https://www.fao.org/faostat/en/#data/QCL>
- Fernandes, L. D. S., Correa, F. M., Ingram, K. T., Almeida, A.-A. F. d., & Royaert, S. (2020). Qtl mapping and identification of SNP-haplotypes affecting yield components of *Theobroma cacao* L. *Horticulture Research*, *7*, 26. <https://doi.org/10.1038/s41438-020-0250-3>
- Goenaga, R., Irizarry, H., & Irish, B. (2009). TARS series of cacao germplasm selections. *HortScience*, *44*(3), 826–827. <https://doi.org/10.21273/HORTSCI.44.3.826>
- Gutiérrez, O., Campbell, A., & Phillips-Mora, W [Wilbert]. (2016). Breeding for disease resistance in cacao. In B. A. Bailey & L. W. Meinhardt (Eds.), *Cacao Diseases* (pp. 567–609). Springer International Publishing. https://doi.org/10.1007/978-3-319-24789-2_18
- Gutiérrez, O., Martínez, K., Zhang, D., Livingstone, D., Turnbull, C. J., & Motamayor, J.C (2021). Selecting SNP markers reflecting population origin for cacao (*Theobroma cacao* L.) germplasm identification. *Beverage Plant Research*, *1*(1), 1–9. <https://doi.org/10.48130/BPR-2021-0015>
- Gutiérrez, O., Puig, A., Phillips-Mora, W [W], Bailey, B. A., Ali, S. S., Mockaitis, K., Schnell, R. J., Livingstone, D., Mustiga, G., Royaert, S., & Motamayor, J. C. (2021). SNP markers associated with resistance to frosty pod and black pod rot diseases in an F1 population of *Theobroma cacao* L. *Tree Genetics & Genomes*, *17*(3), 106. <https://doi.org/10.1007/s11295-021-01507-w>
- ICCO. (2021). *Production of cocoa beans: Quarterly bulletin of cocoa statistics, Vol. XLVII, No. 4, Cocoa year 2020/21*. Côte d'Ivoire. The International Cocoa Organization (ICCO).
- ICCO. (2022). *Fine or Flavour Cocoa*. <https://www.icco.org/fine-or-flavor-cocoa/>
- Lanaud, C., Fouet, O., Legavre, T., Lopes, U., Sounigo, O., Eyango, M. C., Mermaz, B., da Silva, M. R., Llor Solórzano, R. G., Argout, X., Gyapay, G., Ebaiarrey, H. E., Colonges, K., Sanier, C., Rivallan, R., Mastin, G., Cryer, N., Boccara, M., Verdeil, J.-L., . . . Clément, D. (2017). Deciphering the *Theobroma cacao* self-incompatibility system: From genomics to diagnostic markers for self-compatibility. *Journal of Experimental Botany*, *68*(17), 4775–4790. <https://doi.org/10.1093/jxb/erx293>
- Lindt & Sprüngli. (2018). *Annual Report 2017*. <https://www.statista.com/statistics/819288/worldwide-chocolate-consumption-by-country/>

- Livingstone, D., Royaert, S., Stack, C., Mockaitis, K., May, G., Farmer, A., Saski, C., Schnell, R., Kuhn, D., & Motamayor, J. C (2015). Making a chocolate chip: Development and evaluation of a 6K SNP array for *Theobroma cacao*. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 22(4), 279–291. <https://doi.org/10.1093/dnares/dsv009>
- Livingstone, D., Stack, C., Mustiga, G. M., Rodezno, D. C., Suarez, C., Amores, F., Feltus, F. A., Mockaitis, K., Cornejo, O. E., & Motamayor, J. C (2017). A larger chocolate chip-development of a 15K *Theobroma cacao* L. Snp array to create high-density linkage maps. *Frontiers in Plant Science*, 8, 2008. <https://doi.org/10.3389/fpls.2017.02008>
- Lopes, U. V., Pires, J. L., Gramacho, K. P., & Grattapaglia, D. (2022). Genome-wide SNP genotyping as a simple and practical tool to accelerate the development of inbred lines in outbred tree species: An example in cacao (*Theobroma cacao* L.). *PLoS One*, 17(10), e0270437. <https://doi.org/10.1371/journal.pone.0270437>
- Marelli, J.-P., Guest, D. I., Bailey, B. A., Evans, H. C., Brown, J. K., Junaid, M., Barreto, R. W., Lisboa, D. O., & Puig, A. S. (2019). Chocolate under threat from old and new cacao diseases. *Phytopathology*, 109(8), 1331–1343. <https://doi.org/10.1094/PHYTO-12-18-0477-RVW>
- McElroy, M. S., Navarro, A. J. R., Mustiga, G., Stack, C., Gezan, S., Peña, G., Sarabia, W., Saquicela, D., Sotomayor, I., Douglas, G. M., Migicovsky, Z., Amores, F., Tarqui, O., Myles, S., & Motamayor, J. C (2018). Prediction of cacao (*Theobroma cacao*) resistance to *Moniliophthora* spp. Diseases via genome-wide association analysis and genomic selection. *Frontiers in Plant Science*, 9, 343. <https://doi.org/10.3389/fpls.2018.00343>
- Morrissey, J., Stack, J. C., Valls, R., & Motamayor, J. C (2019). Low-cost assembly of a cacao crop genome is able to resolve complex heterozygous bubbles. *Horticulture Research*, 6, 44. <https://doi.org/10.1038/s41438-019-0125-7>
- Motamayor, J. C., Lachenaud, P., da Silva E Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PLoS One*, 3(10), e3311. <https://doi.org/10.1371/journal.pone.0003311>
- Motamayor, J. C., Mockaitis, K., Schmutz, J., Haiminen, N., Livingstone, D., Cornejo, O., Findley, S. D., Zheng, P., Utro, F., Royaert, S., Saski, C., Jenkins, J., Podicheti, R., Zhao, M., Scheffler, B. E., Stack, J. C [Joseph C.], Feltus, F. A., Mustiga, G. M., Amores, F., . . . Kuhn, D. N. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biology*, 14(6), r53. <https://doi.org/10.1186/gb-2013-14-6-r53>
- Nieves-Orduña, H. E., Krutovsky, K. V., & Gailing, O. (2023). Geographic distribution, conservation, and genomic resources of cacao *Theobroma cacao* L. *Crop Science*, 63(4), 1750–1778. <https://doi.org/10.1002/csc.2.20959>
- Phillips-Mora, W., Arciniegas-Leal, A., Mata-Quirós, A., & Motamayor, J. C [J C]. (2013). *Catalogue of cacao clones: Selected by CATIE for commercial plantings. Technical series, technical manual: Nr. 105*. CATIE.
- Romero-Navarro, J. A., Phillips-Mora, W., Arciniegas-Leal, A., Mata-Quirós, A., Haiminen, N., Mustiga, G., Livingstone, D., van Bakel, H., Kuhn, D. N., Parida, L., Kasarskis, A., & Motamayor, J. C. (2017). Application of genome wide association and

- genomic prediction for improvement of cacao productivity and resistance to black and frosty pod diseases. *Frontiers in Plant Science*, 8, 1905.
- Royaert, S., Jansen, J., Silva, D. V., Jesus Branco, S. M. de, Livingstone, D. S., Mustiga, G., Marelli, J.-P., Araújo, I. S., Correa, R. X., & Motamayor, J. C. (2016). Identification of candidate genes involved in Witches' broom disease resistance in a segregating mapping population of *Theobroma cacao* L. In Brazil. *BMC Genomics*, 17, 107.
- Smith, N. J., & Schultes, R. E. (1990). Deforestation and shrinking crop gene-pools in Amazonia. *Environmental Conservation*, 17(3), 227–234. <https://doi.org/10.1017/S0376892900032367>
- Soares-Filho, B. S., Nepstad, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C. A., Voll, E., McDonald, A., Lefebvre, P., & Schlesinger, P. (2006). Modelling conservation in the Amazon basin. *Nature*, 440(7083), 520–523. <https://doi.org/10.1038/nature04389>
- Steege, H. T., Pitman, N. C. A., Killeen, T. J., Laurance, W. F., Peres, C. A., Guevara, J. E., Salomão, R. P., Castilho, C. V., Amaral, I. L., Almeida Matos, F. D. de, Souza Coelho, L. de, Magnusson, W. E., Phillips, O. L., Andrade Lima Filho, D. de, Jesus Veiga Carim, M. de, Ireme, M. V., Martins, M. P., Molino, J.-F., Sabatier, D., . . . Gamarra, L. V. (2015). Estimating the global conservation status of more than 15,000 Amazonian tree species. *Science Advances*, 1(10), e1500936. <https://doi.org/10.1126/sciadv.1500936>
- Steege, H. T., Pitman, N., Sabatier, D., Castellanos, H., van der Hout, P., Daly, D. C., Silveira, M., Phillips, O., Vasquez, R., van Andel, T., Duivenvoorden, J., Oliveira, A. A. de [Alexandre Adalardo], Ek, R., Lilwah, R., Thomas, R., van Essen, J., Baider, C., Maas, P., Mori, S., . . . Morawetz, W. (2003). A spatial model of tree α -diversity and tree density for the Amazon. *Biodiversity and Conservation*, 12(11), 2255–2277. <https://doi.org/10.1023/A:1024593414624>
- Toxopeus, H. (1969). Cacao (*Theobroma cacao* L.). In F. Ferwerda & F. Wit (Eds.), *Outlines of perennial crop breeding in the tropics*. Veenman.
- Tuomisto, H., Ruokolainen, K., Kalliola, R., Linna, A., Danjoy, W., & Rodriguez, Z. (1995). Dissecting amazonian biodiversity. *Science (New York, N.Y.)*, 269(5220), 63–66. <https://doi.org/10.1126/science.269.5220.63>
- Venturieri, G. A. (2013). Indigenous strategies used to domesticate plants in Brazilian Amazon. In *Encyclopedia of Biodiversity* (pp. 279–292). Elsevier. <https://doi.org/10.1016/B978-0-12-384719-5.00302-6>
- Wood, G., & Lass, R. A. (1985). *Cocoa* (Fourth). Blackwell Science Ltd. <https://doi.org/10.1002/9780470698983>
- Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I., Fouet, O., Argout, X [Xavier], Guichoux, E., Salin, F., Solorzano, R. L., Bouchez, O., Vignes, H., Severts, P., Hurtado, J., Yopez, A., Grivetti, L., Blake, M., & Valdez, F. (2018). The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nature Ecology & Evolution*, 2(12), 1879–1888. <https://doi.org/10.1038/s41559-018-0697-x>
- Zhang, D., Figueira, A., Motilal, L., Lachenaud, P., & Meinhardt, L. W. (2011). *Theobroma*. In C. Kole (Ed.), *Wild Crop Relatives: Genomic and Breeding Resources* (Vol. 6, pp. 277–296). Springer. https://doi.org/10.1007/978-3-642-21201-7_13

- Zhang, D., Martínez, W. J., Johnson, E. S., Somarriba, E., Phillips-Mora, W., Astorga, C., Mischke, S., & Meinhardt, L. W. (2012). Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. *Genetic Resources and Crop Evolution*, 59(2), 239–252. <https://doi.org/10.1007/s10722-011-9680-y>
- Zhang, D., & Motilal, L. (2016). Origin, dispersal, and current global distribution of cacao genetic diversity. In B. A. Bailey & L. W. Meinhardt (Eds.), *Cacao Diseases* (pp. 3–31). Springer International Publishing. https://doi.org/10.1007/978-3-319-24789-2_1

Paper 1

Geographic distribution, conservation, and genomic resources of cacao

***Theobroma cacao* L**

Published

Crop Science

22 March 2023

<https://doi.org/10.1002/csc2.20959>

Geographic distribution, conservation, and genomic resources of cacao *Theobroma cacao* L

Helmuth E. Nieves-Orduña^{1,2}, Konstantin V. Krutovsky^{1,2}, Oliver Gailing^{1,2}

Corresponding authors:

Helmuth E. Nieves-Orduña and Oliver Gailing, Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, 37077 Göttingen, Germany.

Email: hnieves@gwdg.de, ogailin@gwdg.de

¹Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, Göttingen, Germany

²Center for Integrated Breeding Research, Georg-August University of Göttingen, Göttingen, Germany

Core Ideas

- The geographic distribution of genetic variation of wild cacao pinpoints new sampling areas for germplasm collections.
- The natural habitat of cacao is under threat due to current and projected deforestation in Amazonia.
- Germplasm collections can benefit from genotype identity and characterization of agronomic traits based on available SNP markers.
- The wide cacao genetic diversity can be tested against different fungal strains to identify new sources of disease resistance.
- Genome-wide SNP data in combination with standard phenotyping of agronomic traits can be exploited to identify relevant genotypes for breeding.

Abstract

The cacao tree, *Theobroma cacao* L., is cultivated in the tropics, mainly in agroforestry systems, to produce seeds, the valuable raw material for the chocolate industry. Thus, the conservation and use of cacao genetic resources in breeding programs to increase yield and improve quality and disease resistance are vital for the global cacao economy. We review three important topics of cacao genetics essential for sustainable production and crop improvement: (1) geographic distribution of wild cacao populations and geographic patterns of their genetic variation in Amazonia; (2) conservation, availability, and use of cacao genetic resources in cultivar development; (3) genetic basis of agronomic traits, available molecular genetic markers, and genomic resources and their application for cacao improvement. We also highlight critical research areas needed to achieve sustainable cacao cultivation.

1 INTRODUCTION

The cacao tree (*Theobroma cacao* L.) is a diploid ($2n = 2x = 20$) species with a comparatively small genome (430–445 Mbp) (Argout et al., 2011; Motamayor et al., 2013). It is a relatively small tree, up to 20 m high in natural forests, and shade-tolerant (Wood & Lass, 1985). Cacao is native to Amazonia and belongs to the Malvaceae family (Bayer & Kubitzki, 2003; Motamayor et al., 2002). The genus *Theobroma* includes 22 species native to tropical America, but none of these wild relatives has contributed to cacao improvement because of strong interspecific crossing barriers (Cuatrecasas, 1964; Zhang et al., 2011). The small and bisexual cacao flowers (developed in small inflorescences) are pollinated by midges of the Ceratopogonidae family (Young, 1994). The cauliflorous cacao fruits of globose to fusiform shape have a hard and rugose surface and exhibit a thick pericarp. On average, there are around 20–40 oblong seeds per fruit (pod) with white and sweet surrounding pulp (Cuatrecasas, 1964).

Cacao is cultivated in 61 tropical countries for the production of seeds, the raw material that shapes the world's chocolate industry (Coe & Coe, 2013; Dand, 2011; FAOSTAT, 2020). Per capita chocolate consumption is very high in both Europe (4–8 kg/year) and the USA (4 kg/year) (Lindt & Sprüngli, 2018). As a tree crop common to small farms and agroforestry systems, cacao seeds are commodities produced primarily in West Africa (76%), followed by Latin America (17%) and Asia (6%) (ICCO, 2021; Nair, 2021). The world's cacao production has been concentrated in West Africa since the 1980s, and top producing countries like Cote d'Ivoire and Ghana have experienced induced deforestation linked to cacao cultivation (Kroeger et al., 2017; Ruf & Schroth, 2004; Wessel & Quist-Wessel, 2015).

Globally, around 2–3 million hectares of tropical forest were lost due to cacao cultivation over a 20-year period (1988–2008) (Kroeger et al., 2017). Low yields, inadequate maintenance, pests and diseases, and increased international demand for chocolate are the main drivers of land use expansion and deforestation linked to cacao cultivation (Hoang & Kanemoto, 2021;

Kroeger et al., 2017; Morrissey et al., 2019; Wessel & Quist-Wessel, 2015). In addition, in West Africa, climate change scenarios predict a rise of 2°C during dry seasons by 2050, reducing suitable lands for cacao cultivation. These potential changes increase the risk of future deforestation in Liberia, Cameroon, and the Congo basin because cacao cultivation may be relocated to cooler areas (Schroth et al., 2016, 2017).

Cacao cultivation may play a dual role in Latin America. For example, it has contributed to an initial deforestation of the Atlantic Forest landscape but to its restoration later in the state of Bahia, Brazil (Bright & Sarin, 2004; Ruf & Schroth, 2004; Schroth et al., 2011). Similarly, both agroforestry reforestation and deforestation were linked to cacao farming in specific areas of Peru and Nicaragua (Orozco-Aguilar et al., 2021). Cacao farming was not associated with deforestation in Colombia but rather opposite—with agroforestry reforestation, which actually is an attractive alternative to illicit crops (Abbott et al., 2018; Barrera-Ramírez et al., 2019; Castro-Nunez et al., 2020; Hernández-Núñez et al., 2022).

As a shade-tolerant tree, more productive cacao can be planted together with valuable native trees creating an agroforestry that increases biodiversity by providing habitat to different plant and animal species (Bhagwat et al., 2008; Blaser et al., 2018; Jezeer et al., 2017; Rice & Greenberg, 2000; Schroth & Harvey, 2007). In addition, cacao agroforestry provides ecosystem services such as carbon capture and storage, water regulation, climate adaptation, and production of timber from shade trees (Blaser et al., 2018; Heming et al., 2022; Martin et al., 2020; Tschardt et al., 2011). Compared to monoculture, cacao agroforestry with an intermediate level of shade (30%–50%) can equally be profitable (Blaser et al., 2018; Jezeer et al., 2017). Market recognition of biodiversity conservation values generated by cacao agroforestry can be achieved through certification schemes (Tschardt et al., 2015).

Due to its importance in national and international markets, the genetic diversity of cacao trees has been studied to explore and improve the crop potential in agronomic traits, such as disease resistance, yield, climate resilience, and seed quality (Bekele & Phillips-Mora, 2019; Bennett, 2003; Daymond & Bekele, 2022; Dias, 2001; Gutiérrez et al., 2016; Hunter, 1990; Schnell et al., 2005). Here, we focus on the conservation and use of cacao genetic resources and the needs to be addressed in the face of current and future challenges.

2 GEOGRAPHIC DISTRIBUTION OF NATURAL POPULATIONS

The wide geographic distribution of cacao in the Amazon basin includes the tropical forests of Peru, Ecuador, Colombia, Bolivia, Brazil, Venezuela, and French Guyana (Figure 1). Cacao is observed from west to east Amazonia in non-flooded and seasonally flooded forests (e.g., high várzea forest, where the water level is less than 3 m in height and lasts less than 50 days/year) (Draper et al., 2019; Wittmann et al., 2006). A recent survey of Amazonia tree diversity observed cacao mostly in non-flooded forests (Steege et al., 2013, 2020). Groups of trees occur in neutral to alluvial soils also scattered and isolated in the understory of primary forests with a density of 1–10 individuals per ha (Allen, 1988; Cuatrecasas, 1964). More recent observations in Madre de Dios, Peru and Rondônia, Brazil showed higher densities, up to 85 and 142 trees/ha, respectively (Steege, 2021). These occurrences may have been influenced by patterns of human introductions in modified forests (Levis et al., 2017; Somarriba & Lachenaud, 2013).

The cacao tree occurs in areas with a mean annual temperature range of 20–30°C and precipitation ranging from 2000 to 6000 mm/year (Figure 2), at an altitude below 1000 m above sea level (masl) (Allen, 1988; Somarriba & Lachenaud, 2013). Recent observations in the South of Peru showed that cacao tolerates dry seasons and can grow at higher altitudes up to 1400 masl (Ceccarelli et al., 2021).

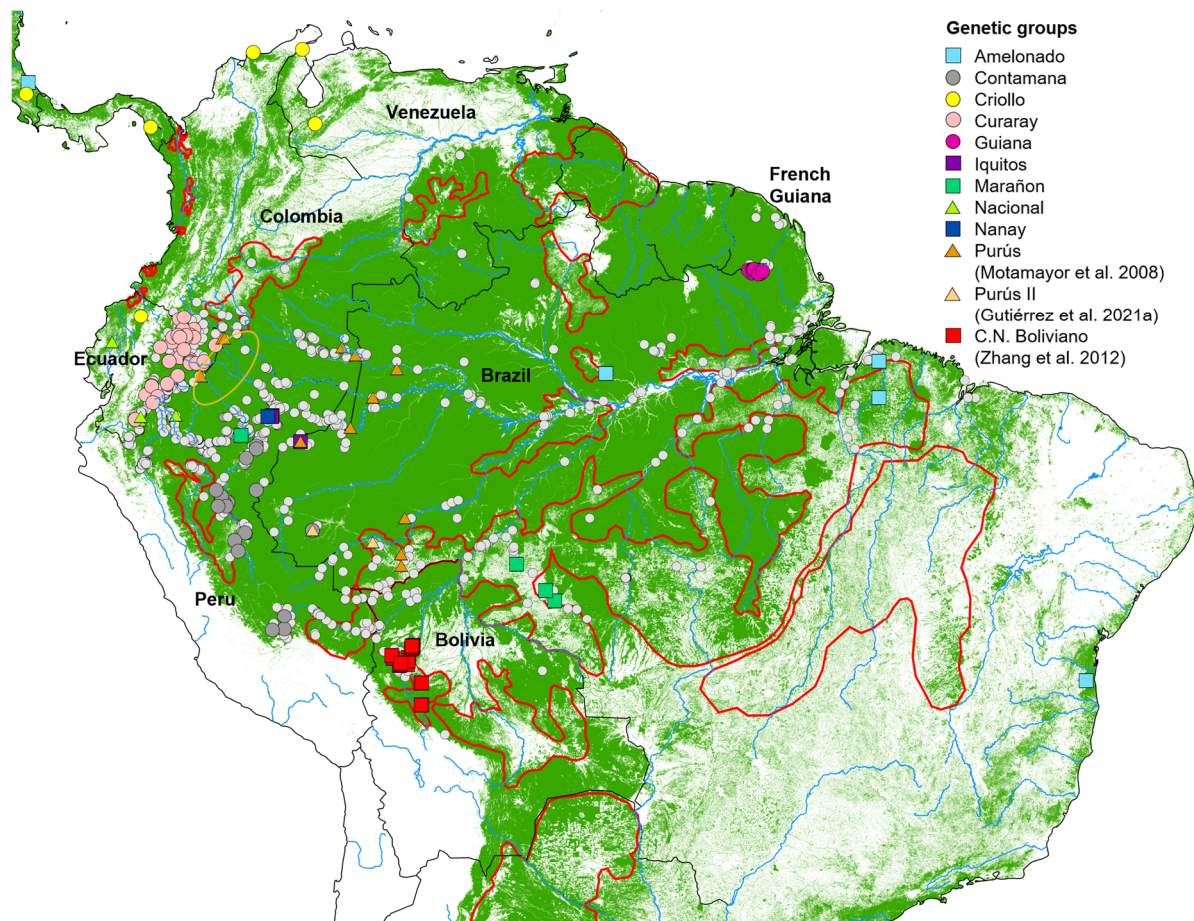


FIGURE 1. Potential geographic distribution of *Theobroma cacao* L., genetic clusters, and habitat loss observed in Amazonia. Light gray circles represent 4413 observations of cacao in wild habitats according to ecological studies and expedition reports for the 1938–2022 time period (Allen, 1988; Almeida, 1983; Almeida et al., 1987, 1995, 2015; Arevalo-Gardini et al., 2019; Barriga, 1982; Carletto, 1973; Ceccarelli et al., 2021; Chalmers, 1968, 1969, 1972, 1973; Chumacero de Schawe et al., 2013; Clement, 1986; Cocoa Research Unit, 1995; Coral, 1988; Desrosiers & Buchwald, 1961; Dias et al., 2003; Fouet et al., 2022; Garcia-Davila et al., 2020; González-Orozco et al., 2020; Lachenaud et al., 1997, 2016; Lachenaud & Sallée, 1993; Lanaud, 1986; Loor et al., 2015; Marita et al., 2001; Morales & Rodriguez, 1987; Motamayor et al., 2008; Pound, 1938, 1943; Sallée, 1987; Sanchez et al., 1988; Sereno et al., 2006; Soria, 1970; Steege, 2021; Turnbull & Hadley, 2021; Vásquez-García et al., 2022; Vello & Medeiros, 1966; Wittmann et al., 2006; Zhang et al., 2012). Observation records include data managed by the Amazon Tree Diversity Network (Steege, 2021) and the International Cocoa Germplasm Database (Turnbull & Hadley, 2021). Cacao genetic clusters are depicted by their respective symbols (circles, triangles, and squares of different colors) according to Motamayor et al. (2008), Zhang et al. (2012), and Gutiérrez, Martínez et al. (2021). The deep green layer shows forest cover (Karra et al., 2021), red polygons represent areas of increased deforestation (Pacheco et al., 2021), and the orange oval shows the border among Ecuador, Colombia, and Peru where cacao observation is low but expected diversity is high (Zarrillo et al., 2018). The map was generated using the ArcGIS software (www.esri.com).

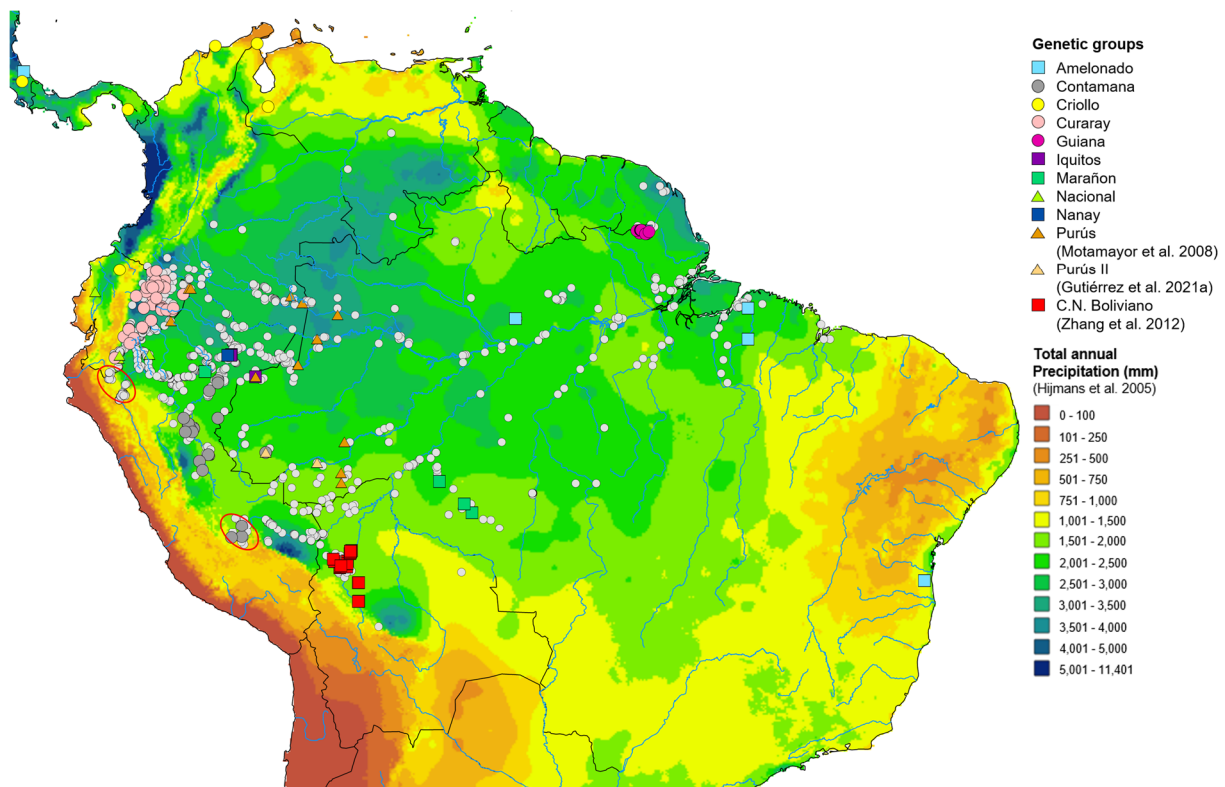


FIGURE 2. Total annual precipitation (Hijmans et al., 2005), geographic distribution of cacao in wild habitats (light gray circles), and genetic groups of *Theobroma cacao* L. in Amazonia. Red ovals show the border between Ecuador and Peru, and the south of Peru (the Contamana group) where cacao grows naturally below 1000 mm of annual precipitation; two locations in Bolivia show a similar pattern. These populations can potentially be used for evaluating drought tolerance traits in cacao breeding. The map was generated using the ArcGIS software (www.esri.com).

2.1 Classification and primary genetic clusters

Morris (1882) (cited in Cuatrecasas, 1964) proposed a classification of cacao populations based on two morphological traits: color of fruits (red to yellow) and fruit surface (smooth or rough). Cacao was divided into two groups: Criollo (red fruits) and Forastero (consisting of eight types, yellow to red fruits and smooth and rough skins) (Morris, 1882). In general, Criollo is known for its smooth fruit skin and low seed productivity, and Forastero is characterized by a rough fruit surface and higher seed yield (Cuatrecasas, 1964). A third group, called Trinitario (Cheesman, 1944), a hybrid (Criollo \times Forastero) originated in Trinidad and Tobago as a result of natural hybridization in cacao farms (Bartley, 2005; Yang et al., 2013). The classification as Criollo, Forastero, and Trinitario groups was traditionally used as a reference of geographical origin of the cacao populations: Forastero from the Amazon basin, Criollo from Central and

South America, and Trinitario from Trinidad, respectively (Bartley, 2005; Cuatrecasas, 1964). However, this initial classification does not reflect the rich genetic diversity of cacao populations observed in Amazonia, making the term Forastero obsolete (Almeida et al., 2015; Arevalo-Gardini et al., 2019; Cornejo et al., 2018; Loor et al., 2015; Motamayor et al., 2008; Motamayor et al., 2010; Zhang et al., 2009).

Motamayor et al. (2008) identified 10 clusters based on the analysis of 952 cacao accessions of widespread geographic origin conserved in *ex situ* collections with 96 simple sequence repeat (SSR) markers. The 10 clusters were named according to the geographic origin of accessions and their genetic clustering based on the SSR markers: Marañón, Curaray, Criollo, Iquitos, Nanay, Contamana, Amelonado, Purús, Nacional, and Guiana (Figures 1 and 2). Three clusters, Criollo, Nacional, and Amelonado, correspond to traditional cultivars. Morphologically, Criollo has plump seeds and white cotyledons; the remaining clusters, which can be grouped as Forastero, have flat seeds and purple cotyledons (Daymond & Bekele, 2022). Hence, this new classification based on cluster analysis of SSR genotypes provides more information about the available genetic resources for conservation and breeding programs (Laliberté et al., 2018; Motamayor et al., 2008; Zhang & Motilal, 2016). The SSR genotypes of 57 wild samples collected along the Beni River in Bolivia revealed an additional cluster, Cacao Nacional Boliviano (Zhang et al., 2012).

Recently, genotypes based on 219 single-nucleotide polymorphisms (SNPs) validated the SSR-based clusters observed in Amazonia and provided higher resolution to study populations and subpopulations within the primary gene pool of cacao in Amazonia (Gutiérrez, Martínez, et al., 2021). For example, accessions from the Napo River in Ecuador assigned to the Purús cluster with a low support in the SSR study were firmly assigned to this cluster based on the SNP genotypes. Moreover, samples geographically restricted to the Purús River in Brazil formed a new cluster—population Purús II—separate from the original Purús cluster. In

addition, the Iquitos cluster is likely restricted to the Peruvian Amazon based on the SNP genotypes, excluding samples collected in Central Amazonia (Gutiérrez, Martínez, et al., 2021). Thus, the 219 SNP panel can be used to classify cacao germplasm more accurately. Overall, SNP genotypes from accessions conserved in national cacao collections may reveal unique ancestry based on geographic origin.

Collecting and analyzing different wild populations in Amazonia with total estimated number of 180 million cacao trees (Steege et al., 2020) might reveal novel cacao genetic diversity and clusters (Bartley, 2005; Laliberté et al., 2018; Zhang et al., 2009; Zhang & Motilal, 2016). For example, wild cacao was observed in northern Amazonia (Figure 1) (Colombia, Venezuela, and Brazil), but it was not studied yet, and no genetic clusters were reported for these areas. Similarly, important river systems in northwestern Amazonia like the upper Napo, Tigre, Pastaza, and Morona in Peruvian Amazon could host new genetic diversity (area depicted by the orange oval in Figure 1), especially considering the frequent observation of wild cacao and occurrence of five genetic clusters in the upper Amazon (Figure 1).

2.2 Geographic origin of clusters

The geographic origin of cacao genetic diversity in Amazonia is attributed to paleoclimate and paleoarches (Motamayor et al., 2008). The paleoarches were ridges no longer visible in the landscape that shaped the genetic structure of cacao by acting as gene flow barriers between cacao clusters (Motamayor et al., 2008). The paleoclimate refers to past dry and cool conditions during the Quaternary period in Amazonia (van der Hammen & Hooghiemstra, 2000). According to the refugia theory, forests were reduced to islands or refugial areas, where temperature, humidity, and edaphic conditions protected tree growth and diversity and promoted genetic differentiation between populations due to geographic isolation (Gentry, 1982; Haffer & Prance, 2001; Prance, 1973). In addition, present-day adaptation of tree species to gradients of soil fertility in Amazonia also explains the occurrence of tree populations in

specific habitats with interrupted gene flow (Gentry, 1989). In cacao, the geographic distribution of genetic diversity is not only attributed to a positive response to fertile soils but also to the geographic isolation of populations growing along different Amazonian rivers system, where gene flow among populations is limited, favoring genetic differentiation (Almeida, 2001; Zhang & Motilal, 2016).

Based on the rich diversity of crop species observed in Amazonia and putative forest refugia, a center of crop genetic diversity for fruit tree species, including cacao, was proposed in western Amazonia (Clement, 1989, 1999). Lachenaud (1997) suggested the existence of forest refugia during the Pleistocene epoch to explain the actual heterogeneity and discontinuous distribution of wild cacao in Amazonia. Likewise, five forest refugia for cacao in Amazonia were proposed by modeling potential past geographic distributions to explain the putative origin of cacao clusters (Thomas et al., 2012). The rich chloroplast diversity observed in wild cacao compared to cultivated cacao supports the forest refugia theory in north western Amazonia (Nieves-Orduña et al., 2021). Furthermore, the distribution of cacao genetic diversity in Amazonia may have also been affected by ancient and modern human intervention (Levis et al., 2017; Peters, 2000; Somarriba & Lachenaud, 2013; Thomas et al., 2012), which may explain the convergence of five genetic clusters (Iquitos, Nanay, Marañon, Contamana, and Purús) in the Peruvian Amazon, and the patterns of diversity decline from west to east in Amazonia (Figure 1; Cornejo et al., 2018; Motamayor et al., 2008; Nieves-Orduña et al., 2021).

3 EX SITU AND IN SITU CONSERVATION OF CACAO

3.1 Cacao expeditions and germplasm collections

After the collapse of cacao plantations in Trinidad around 1930 as a result of witches' broom disease (WBD) caused by *Moniliophthora perniciosa*, there was a need to look for new sources of genetic resistance in natural populations of Amazonia (Bartley, 2005). The narrow genetic

base of cultivated cacao also became evident later on in Bahia, Brazil, where in the 1990s the production decreased dramatically due to the vulnerability to WBD (Bennett, 2003). Thus, cacao germplasm collections in Amazonia have been motivated by the search for disease resistance and yield traits (Bartley, 2005; Zhang et al., 2011). The search for disease resistant genotypes focused extensively on the northern Peruvian Amazon by the collector F.J. Pound during the 1930–1940s (Bartley, 2005; Zhang et al., 2011). More recently, systematic field trips aiming to collect genotypes expressing disease resistance and superior yield traits searched wild cacao trees in the Colombian and Ecuadorian Amazon (Allen, 1988). Detailed descriptions of historical cacao expeditions collecting material for *ex situ* collections are presented in Zhang et al. (2011), Zhang and Motilal (2016), and Rodriguez-Medina et al. (2019).

New cacao expeditions were motivated to widen the genetic base for cacao breeding and filling gaps in *ex situ* collections (Bekele & Phillips-Mora, 2019; Daymond & Bekele, 2022; Laliberté et al., 2018). Collection trips aiming to capture new genetic diversity for disease resistance and quality traits were done in Peru (Arevalo-Gardini et al., 2019; Vásquez-García et al., 2022; Zhang & Motilal, 2016), Ecuador (Fouet et al., 2022; Loor et al., 2015), and French Guiana (Lachenaud et al., 2018). In addition, not only disease resistance but also quality traits have been subject of specific cacao collections. In northern Peru, collections of wild cacao were conducted to investigate populations of the fine flavored cacao variety “Piura Porcelana” (Arevalo-Gardini et al., 2019). Specific collections aiming to rescue and conserve wild ancestors of the fine flavored Nacional group were conducted in the Ecuadorian Amazon (Fouet et al., 2022; Loor et al., 2015). In addition, expeditions in the Colombian Amazon along the Caquetá and Caguan rivers identified wild cacao populations and potential areas for germplasm collection (González-Orozco et al., 2020; Rodriguez-Medina et al., 2019).

New germplasm collections could focus on areas where expected genetic diversity is high (Figure 1), such as northwestern Amazon (Motamayor et al., 2008; Nieves-Orduña et al., 2021; Thomas et al., 2012; Zarrillo et al., 2018). However, germplasm collections have to be done on time before deforestation and climate change erodes the natural habitat of cacao in Amazonia (Gomes et al., 2019; Soares-Filho et al., 2006).

3.2 *Ex situ* conservation

Cacao needs to be conserved as a living plant. Due to the recalcitrant nature of cacao seeds, they can be conserved only up to 10 weeks and do not tolerate desiccation below 27% and temperatures lower than 17°C (Hor et al., 1984). Hence, cacao needs to be conserved as living collections of trees, thus each tree represents a unique genotype in germplasm collections (CacaoNet, 2012). The *ex situ* conservation of cacao genetic diversity is currently arranged mainly via 47 national collections and two international collections under the public domain regulated by the International Treaty on Plant Genetic Resources for Food and Agriculture (Motilal, 2018; Thormann et al., 2015). International collections are maintained in Costa Rica at CATIE (International Center for Tropical Agriculture) and Trinidad and Tobago at ICGT (International Cocoa Genebank Trinidad, West Indies University). Accessions maintained in CATIE and ICGT collections represent wild, breeding populations (mostly as admixture trees), and genetic clusters. Among the genetic clusters conserved, Marañón, Nanay, Iquitos, and Amelonado are the most frequent, Criollo is the least represented, and Cacao Nacional Boliviano is absent.

A core germplasm collection aims to capture the total crop genetic diversity by including an optimal set of accessions for trait evaluation and providing access for breeders (Maxted et al., 2020). Two hundred sixty-one germplasm accessions were proposed for a cacao core collection, including the 10 genetic clusters identified by Motamayor et al. (2008) plus 11 Trinitario accessions (CacaoNet, 2012; Motilal, 2018). The criteria for germplasm selection

did not include agronomical or morphological trait information, but individual SSR and geographic genetic variation data observed in different *ex situ* collections (CacaoNet, 2012). In the case of Guianan cacao, through the analysis of 1953 SNPs, 41 out of 181 cacao clones were selected as a core collection, but 21 clones captured 99% of all alleles in a core collection (Lachenaud et al., 2018).

Cacao is conserved also in national germplasm *ex situ* collections mainly in countries with natural cacao populations such as Brazil (4773 accessions), Ecuador (1939), Peru (593), Colombia (411), Venezuela (419), and French Guiana (551) (Turnbull & Hadley, 2021). Among them, Brazil and Ecuador are top cacao producing countries in Latin America (ICCO, 2021). Although cacao is native to Bolivia, there are no records of *ex situ* conservation in this country (Turnbull & Hadley, 2021). The west African countries, which produce 76% of global cacao (ICCO, 2021), also hold cacao accessions, for example, Cote d'Ivoire (411), Ghana (1278), and Cameroon (230) (Turnbull & Hadley, 2021). In total, around 18,634 cacao germplasm accessions are conserved in national *ex situ* collections (Turnbull & Hadley, 2021). When considering specific national regulations, breeders may face challenges to access and exchange cacao accessions due the exclusion of cacao from Annex I of the International Treaty on Plant Genetic Resources for Food and Agriculture. As a response, the Global Strategy for cacao genetic resources aims to improve conservation and access of cacao germplasm through the creation of a network of international partners, including countries that maintain *ex situ* collections, national authorities, and the private sector like chocolate companies (CacaoNet, 2012; Laliberté et al., 2018).

Agronomic evaluation carried out in *ex situ* collections has been fundamental for selecting commercial cacao clones and their incorporation in breeding programs (Bekele & Phillips-Mora, 2019). For example, in Trinidad and Tobago, among 1900 diverse cacao accessions conserved and evaluated at the ICGT, 57 Trinitario clones were recommended for cacao

plantations and as parents in breeding programs for their favorable yield potential, bean quality, and chocolate flavor (Bekele et al., 2020). Trinitario accessions have an impact on cacao producing countries worldwide (Bartley, 2005). Distributed in Latin America, Asia, and Africa, around 36 common Trinitario accessions (e.g., ICS and UF clones) have been used as planting material and breeding accessions (Bekele et al., 2020; Bekele & Phillips-Mora, 2019; Zhang et al., 2011). In French Guiana, CIRAD recommended three clones (known as GU clones) because of their disease resistance against WBD and black pod rot (BPR), yield performance, and intense chocolate flavor (Lachenaud et al., 2007). The agronomic evaluation of GU clones lasted 10 years as an initial population of 1600 wild cacao trees was collected in Guianan forests (Lachenaud et al., 2007).

Similarly, after 15 years of field trials, six clones with high yield and disease resistance to monilia or frosty pod rot (FPR) disease (caused by *Moniliophthora roreri*) and BPR disease were selected in the breeding program at CATIE in Costa Rica (Phillips-Mora et al., 2012). These selected clones (e.g., CATIE R6) have been widely distributed to cacao farmers in Central America and Brazil (Phillips-Mora et al., 2017). In addition, nine superior hybrids (TARS series) were selected after 4 years of evaluating yield and disease resistance in the germplasm collection maintained by the USDA in Puerto Rico (Goenaga et al., 2009). These 9 cultivars originated from a breeding population of 1310 trees representing 5 full-sib families (UF-668 × Pound-7, IMC-67 × UF-613, EET-400 × SCA-12, SCA-6 × EET-62, and IMC-67 × SCA-12) (Goenaga et al., 2009). Each clonal parent was previously selected in CATIE for its high yield and disease resistance traits (Goenaga et al., 2009). Further selections in Puerto Rico yielded three new cultivars (Gasic et al., 2016). Recent agronomic evaluation of native cacao from Peru identified 24 out 103 accessions with potential for the chocolate industry (Vásquez-García et al., 2022).

The cacao germplasm data are stored, and their access is facilitated by the International Cocoa Germplasm Database (ICGD, <http://www.icgd.rdg.ac.uk>) (Turnbull & Hadley, 2021). The ICGD provides valuable primary information on germplasm availability and agronomic traits for the research community, industry, and cacao growers (Daymond & Bekele, 2022). By collecting, updating, and managing germplasm information, the ICGD supports the selection of cacao accessions for conservation and breeding purposes. For example, breeders interested in cacao clones with confirmed resistance to FPR disease can check agronomic data of 18 cacao clones, including geographic origin, pedigree (when available), and the studies that support the selection of these disease resistant clones (Turnbull & Hadley, 2021).

3.2.1 Cryopreservation

Cryopreservation is an alternative to living collections for tree crops with recalcitrant seeds. In vitro material can be conserved at -196°C without losing viability for many plants (O'Brien et al., 2021). Reference germplasm collections can be preserved safely and cost-effectively in liquid nitrogen tanks when standard cryopreservation protocols are available (Jenderek & Reed, 2017). Cryopreservation of cacao somatic embryos was developed as a long-term alternative for genetic conservation (Adu-Gyamfi & Wetten, 2012, 2020; Fang et al., 2004). Cacao cryopreservation relies on floral-derived secondary somatic embryos because there are no methods for micropropagation of cacao shoot tips (Adu-Gyamfi & Wetten, 2020). Cryopreservation has been tested in six cacao genotypes, where the cultivar CCN51 exhibited a lower survival rate of 36% (Adu-Gyamfi & Wetten, 2012). However, to date, there are no cacao cryo-banks as a backup of *ex situ* collections. It seems that more research is needed to optimize and implement cacao cryopreservation protocols (vitrification) for specific genotypes (like cultivars and cacao core germplasm accessions). In addition, cacao embryogenesis, which tends to be genotype dependent, also needs to be validated in accessions of interest for cryopreservation (Wickramasuriya & Dunwell, 2018). Other research topics in developing

cryopreservation protocols should include studies of epigenetic and genetic changes and field performance of plants regenerated after cryopreservation (Wang et al., 2021). In cacao, phenotypic variability was observed in the cryopreserved accession AMAZ 15 and could be due to DNA methylation changes (Adu-Gyamfi et al., 2016).

3.3 *In situ* conservation

The natural habitats that host not only the primary genetic pool of cacao but also other 6000 tree species are under threat because of the continued destruction of Amazonian forests (Cardoso et al., 2017; Gomes et al., 2019; Rojas et al., 2021; Steege et al., 2015; Vancutsem et al., 2021). Unfortunately, it is projected that more than 40% of Amazonia will be lost by 2050 because of agricultural expansion and climate change (Gomes et al., 2019; Soares-Filho et al., 2006). The *in situ* conservation of cacao secured by national protected areas in Amazonia is experiencing different levels of threats due to illegal mining and logging, fires, and agricultural expansion (Armenteras et al., 2021; RAISG, 2020; Soares-Filho et al., 2006). Recently, five areas were proposed for *in situ* conservation and germplasm collection of wild cacao in the Colombian Amazon, and areas for cacao wild relatives (*Theobroma* spp. and *Herrania* spp.) were also identified in this study (González-Orozco et al., 2020).

Conservation strategies for widespread tropical tree species, such as cedar or cedro amargo (*Cedrela odorata*), have been proposed based on the distribution of cpDNA variants (haplotypes) in its natural habitat (Cavers et al., 2004). In the case of cacao, a center of cpDNA diversity was observed in the northwestern Amazon, including the occurrence of seven geographically restricted chloroplast haplotypes in the Peruvian Amazon (Nieves-Orduña et al., 2021). Because cacao is a non-pioneer species, *in situ* conservation needs to consider ecological conditions provided by natural forests to favor seed germination and adaptive changes (Akinagbe et al., 2019; Finkeldey & Hattemer, 2007). However, research projects investigating the *in situ* conservation of cacao are scarce (Daymond & Bekele, 2022). The study

of cacao conservation units depends on the genetic variation present in Amazonia; therefore, an initial step should include population genetic analysis of wild populations conserved in national *ex situ* collections. In addition, novel genetic diversity in natural populations sampled in new collections trips should be studied and used in breeding programs. It is also important to study population structure, mating system, gene flow, and ancestry of cacao populations using selectively neutral cacao markers, such as SSRs and SNPs (Gutiérrez, Martínez, et al., 2021; Mahabir et al., 2020; Motamayor et al., 2008; Wang et al., 2020; Wever et al., 2019).

4 GENETIC BASIS OF AGRONOMIC TRAITS

Cacao breeding for chocolate production is still a challenge that requires the integration of data on the best agronomic traits obtained from different populations with modern genomic and breeding analytical and experimental tools. Figure 3 demonstrates how breeding process can profit from combining standard phenotyping with studies such as genome-wide association studies (GWAS) for determining the genetic basis of economic traits (right part of the figure). Pre-breeding stages include also accurate genotype identification of germplasm collections and selection of agronomic traits based on available SNP markers (left part of the figure). Marker-assisted selection, genomic selection (GS) to predict phenotype performance of new genotypes, and field trials can be used together to validate new genotypes before commercial planting (center part of the figure). Overall, the genetic basis of agronomic traits relies on the conservation and use of cacao genetic diversity. For example, the study of complex traits among different cacao populations (Figure 1) aided with GWAS can identify novel and useful alleles for breeding disease resistant and low cadmium (Cd) genotypes (Figure 3). Breeding programs can obtain pathogen-free cacao plants (typically budwoods) from the International Cocoa Quarantine Center in the United Kingdom (Daymond, 2018). However, access to more diverse populations is often limited due to phytosanitary control and genetic resources' protective regulations in countries with natural cacao populations and *ex situ* collections

(Laliberté et al., 2018). Overall, increasing the number of geographically distant individuals (e.g., up to 1000 in GWAS by da Silva et al., 2016) representing a wide range of phenotypic variation genotyped with a high-density SNP array (e.g., up to 100 K) can reveal novel marker-trait associations.

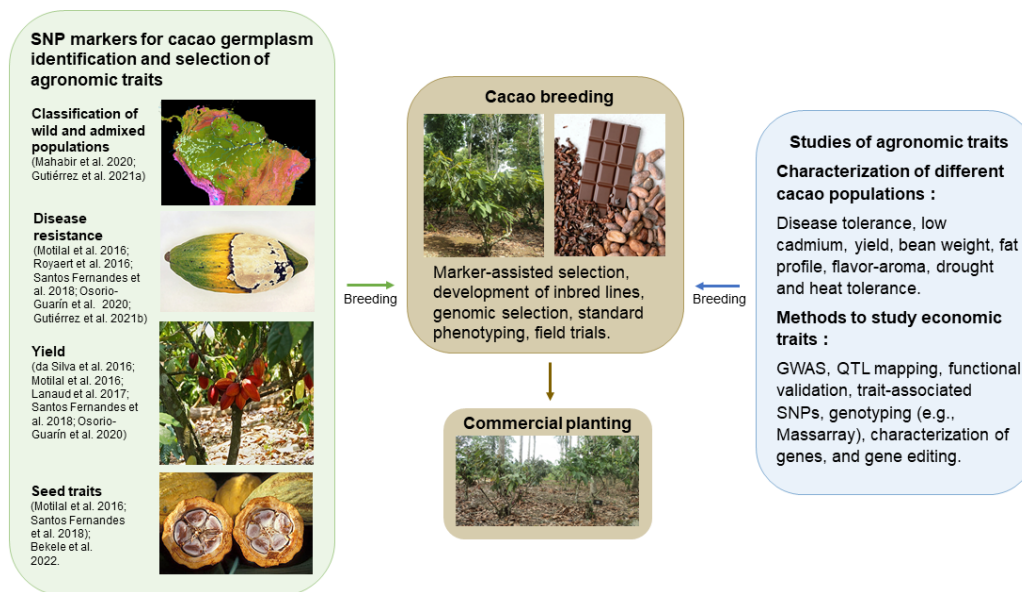


FIGURE 3. Use of genetic markers and modern genomic tools in combination with studies of agronomic traits for cacao breeding. Cacao fruit photos courtesy of USDA ARS.

In addition, standardization of phenotyping of economically important traits is essential in GWAS experiments. For example, the evaluation of FPR disease resistance over 4 years proved that quantitative internal disease measures, such as percentage of necrosis in fruits, were more accurate than external qualitative disease evaluation (Gutierrez et al., 2021). Thus, measurements of internal disease severity are better suited to select FPR resistance genotypes (Gutierrez et al., 2021). In addition, reduced yield variability from the same clone can be achieved by measuring seed weight from fruits manually pollinated or using the number of seeds in the fruits as a covariable (Doaré et al., 2020). Either option will normalize seed weight phenotyping, providing data comparability among breeding programs (Doaré et al., 2020).

Validation of candidate SNPs identified by GWAS under different plantation environments is also needed, although it may represent a challenge in cacao breeding.

4.1 Genetic basis of disease resistance

Diseases are common in cacao and destroy up to 38% of the annual global yield, badly affecting farmers, and the chocolate industry (Bowers et al., 2001; Marelli et al., 2019; Ploetz, 2016); with world market revenue estimated at 980 billion in 2021 (Statista, 2022). The top main cacao diseases affecting cacao cultivation are BPR caused by seven species of *Phytophthora* sp., WBD caused by *M. pernicioso*, cacao swollen shoot virus caused by Badnavirus, and FPR caused by *M. royeri* (Bowers et al., 2001; Gutiérrez et al., 2016; Marelli et al., 2019). The world chocolate economy will be at risk, if FPR and WBD, which currently are geographically restricted to tropical America, would be introduced to West Africa, where more than 70% of the global production relies on disease susceptible Amelonado varieties (Díaz-Valderrama et al., 2020; Marelli et al., 2019; Ploetz, 2016).

4.1.1 Black pod rot disease (BPR)

Globally, BPR is the most economically harmful cacao disease, especially in Africa, where yield losses can be up to 90%, if not controlled (Morales-Cruz et al., 2020). It is caused by two significant pathogens: *Phytophthora palmivora* (native to Southeast Asia) and more aggressive *Phytophthora megakarya* (native to Central Africa and confined to Central and West Africa) (Morales-Cruz et al., 2020). Some cacao natural populations and cultivars show resistance to BPR, as well as disease resistance candidate genes are also identified (Table 1). For example, the Guiana group showed natural resistance to *P. megakarya*: Disease resistance was evaluated using artificial inoculation with *P. megakarya* in detached cacao leaves (Paulin et al., 2008), and the resistance of Guiana progenies to BPR was confirmed in field evaluations in Ghana (Ofori et al., 2020). Field evaluation of 148 individuals with admixed ancestry showed that trees with higher Nacional ancestry were more susceptible to BPR, whereas those with

TABLE 1. Disease resistance candidate genes for prospective genetic improvement of cacao *Theobroma cacao* L.

Gene	Chromosome	Molecular description	Reference	
Associated with resistance to frosty pod rot (FPR) disease caused by <i>Moniliophthora roreri</i>				
TCM_020057	4	trithorax-like protein 2 isoform 2	Romero Navarro et al., 2017	
TCM_020262	4	phosphate transporter 1,5		
TCM_025566	5	gamma-tocopherol methyltransferase		
TCM_025986	5	uncharacterized protein		
TCM_027463	6	kinase superfamily protein		
TCM_045155	10	uncharacterized protein		
Tc00cons_t021470.1	1	S-locus lectin protein kinase family	Osorio-Guarín et al., 2020	
Tc02v2_g007330	2	MLO-like protein		
TCM1EG006844	2	Thymidylate synthase 1	Gutiérrez et al., 2021b	
TCM1EG017087	4	Ankyrin repeat-containing protein		
TCM1EG031605	7	Ribosomal protein L11 methyltransferase-related		
TCM1EG034906	8	Mitochondrial substrate carrier family protein		
TCM1EG042307	9	Phototropic-responsive NPH3 family protein		
TCM1EG041978	9	Phototropin 1		
TCM1EG042792	10	Family not named		
TCM1EG042522	10	protein kinase family protein		
Associated with resistance to witches' broom disease (WBD) caused by <i>M. perniciosa</i>				
Thecc1EG016568	3	RNA-binding (RRM/RBD/RNP motifs) family protein isoform 2		Royaert et al., 2016
Thecc1EG016775	4	UDP-glucose pyrophosphorylase 2		
Thecc1EG028963	6	hydroxyproline-rich glycoprotein family protein, putative		
Thecc1EG030009	6	MYB-like 102		
Thecc1EG032345	7	nucleotide/sugar transporter family protein		
Thecc1EG038262	9	RING/U-box superfamily protein		
Thecc1EG047089	9	RNA recognition motif (RRM)-containing protein		
Tc02v2_g014130	2	G-type lectin S-receptor-like serine/threonine protein	Osorio-Guarín et al., 2020	
Tc02v2_g001090	2	xyloglucan galactosyltransferase		
Tc02v2_g007330	2	MLO-like protein		
Tc03v2_g015970	3	protein IWS1		
Tc06v2_p000370	6	pathogenesis-related (PR)-16	Mourmet et al., 2020	
Tc01v2_g005650	1	Putative Probable receptor-like protein kinase At5g24010	Chia Wong et al., 2022	
Tc01v2_g005810	1	Putative disease resistance protein RGA3		
Tc08v2_g003970	8	Rac-like GTP-binding protein RAC1		
Tc09v2_g027810	9	Putative Probable LRR receptor-like serine/threonine-protein kinase		
Associated with resistance to black pod rot (BPR) disease caused by <i>Phytophthora sp.</i>				
TcNPR3	6	non-expressor of pathogenesis-related 3	Fister et al., 2018	
TCM1EG007680	6	Nucleotidyl transferase superfamily protein	Gutiérrez et al., 2021b	
TCM1EG020430	4	Core-2/I-branching beta-1		
TCM1EG034751	8	NRAMP metal ion transporter 2		
TCM1EG045112	10	Uncharacterized protein		
TcWRKY29	3	WRKY transcription factor 29	Winters, 2022	
TcCSE	6	caffeoyl-shikimate esterase		
TcBBE8	6	berberine bridge enzyme 8		

Marañón ancestry showed resistance to BPR (Romero-Navarro et al., 2017). Furthermore, inoculation of 60 different cacao genotypes identified 22 accessions with resistance to BPR, of

which 4 exhibited high disease resistance in some accessions in Nanay (2 accessions), Iquitos (1 accession), and Marañon (1 accession) genetic groups (Fister et al., 2020).

Using published SSR markers and field response to BPR of traditional cacao varieties (Comum, Pará, and Maranhão) cultivated in Bahia, Brazil, 160 candidate genes were identified as associated with resistance to BPR (Mucherino Muñoz et al., 2021). The candidate genes include RLK (receptor-like kinase domain) and RLK + GNK2 (ginkbilobin-2), coiled-coils, and nucleotide-binding sites. From the breeder's perspective, four SSR markers were associated with BPR resistance, explaining 1.93%–7.43% of the phenotypic variation (Mucherino Muñoz et al., 2021). Gene expression analysis of leaves inoculated with *P. palmivora* revealed higher expression levels of disease resistance genes in the BPR resistant genotype Pound 7 compared to common cacao cultivars ICS 1 and CCN 51 (Baruah et al., 2021). The differentially expressed genes included genes of the KEGG pathway (phenylpropanoid biosynthesis), CAZY families, ubiquitin related proteins, LRR (leucine-rich repeat), and NB-ARC (nucleotide-binding adaptor) (Baruah et al., 2021). Furthermore, gene expression analysis revealed 67 pathogenesis-related (PR) genes when cacao seedlings were infected with *P. palmivora* (Fister et al., 2016). Although the cacao PR gene family did not stop the infection (after 72 h of inoculation, the pathogen was re-isolated to prove successful infection), it demonstrates that PR genes in cacao are involved in leaf tissue defense against BPR (Fister et al., 2016). In addition, a pioneer study in cacao using the CRISPR/Cas9 strategy demonstrated that the inactivation of non-expressor of PR 3 (TcNPR3) gene suppressed cacao disease response, and deletions in the TcNPR3 gene in detached leaves showed more resistance to BPR (Fister et al., 2018). Another gene involved in cacao resistance to BPR includes the caffeoyl-shikimate esterase (CSE) gene, which synthesizes caffeic, an inhibitor of *P. palmivora* (Winters, 2022). Consistent upregulation of TcCSE gene expression in response to *P. palmivora* infection was observed across BPR resistant genotypes of Guiana, Iquitos, Marañon, and Nanay groups

(Winters, 2022). Other BPR disease resistance candidate genes identified across five species of *Theobroma* that also showed upregulated gene expression in response to *P. palmivora* infection were TcWRKY29 (WRKY transcription factor 29) and TcBBE8 (berberine bridge enzyme 8) (Winters, 2022).

These findings provide evidence of natural resistance to BPR among wild cacao populations. Incorporation of disease resistance while including additional traits, such as higher yield in single genotypes, still remains a challenge for breeding programs, but progress has been made in West Africa (Ofori et al., 2020) and Tropical America (Gutiérrez et al., 2016; Gutiérrez, Puig, et al., 2021; Phillips-Mora et al., 2012).

4.1.2 Witches broom disease (WBD)

WBD caused by *M. pernicioso* is one of the most critical cacao diseases, particularly in Brazil, but, fortunately, wild cacao populations exhibit some natural resistance. Restricted to tropical America, and native to Amazonia, the range of hosts of *M. pernicioso* includes *Theobroma* spp. and trees from the Solanaceae family (Lisboa et al., 2020). Cacao yield losses can be very high when WBD occurs together with FPR (Marelli et al., 2019; Ploetz, 2007). Resistance to WBD was observed in wild cacao populations, such as C SUL and SCA accessions, both collected in the upper Amazon and used by breeding programs in Brazil (Bennett, 2003; Marita et al., 2001; Paim et al., 2006; Pimenta Neto et al., 2018). However, because *M. pernicioso* in Brazil could overcome the disease resistance of SCA 6, the primary source of disease resistance, new resistant cultivars are needed to sustain the cacao production (Pereira et al., 2021). Thus, field evaluations of WBD in Brazil using 11 cacao genotypes with wide geographic origin identified the wild cacao accessions C SUL-3 (collected in Western Amazon, Brazil) and GU-171 (collected in French Guyana) as novel sources for the WBD resistance (Pereira et al., 2021).

The genetic basis of cacao resistance to WBD is polygenic (Table 1). The analysis of progeny in the mapping population from the TSH 1188 × CCN 51 cross identified seven quantitative trait loci (QTL) with small effects on the WBD resistance (Royaert et al., 2016). The major QTL on chromosome 9 originated from the SCA 6 accession. Candidate genes were validated through expression data, and SNP markers with highest association with disease resistance were identified in seven candidate genes (Royaert et al., 2016). Recently, WBD resistance was also studied in the F2 population derived from the SCA 6 × ICS1 cross, and 23 QTL with minor effects were identified (Chia Wong et al., 2022). Three identified candidate genes encoding glutathione peroxidases, threonine–serine receptors, and endochitinases, respectively, were associated with WBD resistance (Chia Wong et al., 2022). Cultivated to produce pulp and seeds, the cupuassu tree (*Theobroma grandiflorum*) is also infected by *M. pernicioso*. The WBD disease resistance gene TgPR3 identified in cupuassu is orthologous to the cacao gene Tc06v2_p000370 (Mournet et al., 2020). The latter confers disease resistance against BPD, thus, validating these two genes as part of defense mechanisms against pathogens in two *Theobroma* species (Mournet et al., 2020).

Breeding for disease resistance to WBD should consider the geographic and genotypic variation of *M. pernicioso*, as resistant clones identified in Brazil and Trinidad were susceptible in Ecuador (Meinhardt et al., 2008). Indeed, *M. pernicioso* has a wide natural geographic range in Brazil (Lisboa et al., 2020), and its population structure revealed dispersal patterns from west Amazonia (isolates collected in wild forests) to the eastern cacao plantations of Bahia (Artero et al., 2017). Furthermore, three *M. pernicioso* genetic clusters without clear host specificity were observed in five cacao production areas of Colombia (Jaimes et al., 2022). Thus, the population structure of *M. pernicioso* was shaped by geographic origin and environmental differences rather than by host specificity (Jaimes et al., 2022). Overall, geographically different *M. pernicioso* isolates obtained from wild and cultivated cacao aided

by additional genetic markers (e.g., SSRs and SNPs) will inform about the population structure and distribution of *M. pernicioso* in cacao production regions of tropical America. This insight will be of value to national breeding programs studying host specificity and resistance of cacao populations against different WBD strains.

4.1.3 Frosty pod rot (FPR) disease

FPR caused by *M. royeri* is the most severe cacao disease, which is restricted to tropical America, excluding Brazil. Yield losses due to FPR can be up to 80%, forcing land use change or “boom-bust” cycles (Zhang & Motilal, 2016). It is a specialized pathogen that attacks only *Theobroma* spp. and *Herrania* sp. trees. Until recently *M. royeri* was considered native to the upper Magdalena River in Colombia, outside the natural geographic distribution of cacao (Ali et al., 2015; Phillips-Mora et al., 2007). However, recent population analysis of *M. royeri* based on SSR and SNP markers identified six genetic clusters in areas of intense cultivation in tropical America: Four clusters represented Coastal Ecuador and three—Magdalena Valley in Colombia, and one cluster included both countries (Díaz-Valderrama et al., 2022). Another study based on SSR markers identified high genetic variation in *M. royeri* from the Ecuadorian Amazon where three clusters were identified (Espinoza-Lozano et al., 2022). A comparison of *M. royeri* collected from plantations in coastal Ecuador and a cluster representing Amazonia suggests the introduction of *M. royeri* from the Ecuadorian Amazon to Coastal Ecuador (Espinoza-Lozano et al., 2022).

The putative geographic origin of *M. royeri* is the Upper Amazon, an area shared by Colombia, Ecuador, and Peru (Díaz-Valderrama et al., 2022). Likely, untapped cacao genetic diversity evolved in the presence of different variants of *M. royeri* in the upper Amazon (Figure 1). Indeed, a recent study shows that the accessions EBC 6 and EBC 9, collected in the Colombian Amazon (Allen, 1988), are resistant to FPR in Colombia (Osorio-Guarín et al.,

2020). Further cacao collections should target the upper Amazon (e.g., the border between Ecuador and Peru) to collect cacao genotypes with new genes conferring resistance to FPR.

Cacao natural resistance to FPR has been identified in germplasm collections and plantations, but it is not common (Bailey et al., 2018; Phillips-Mora & Wilkinson, 2007). FPR artificial inoculation of fruits allowed to test the tolerance reaction of diverse cacao accessions against seven FPR isolates (Phillips-Mora et al., 2005). Using this approach on 746 cacao accessions with geographically diverse origin maintained at CATIE revealed limited FPR resistance: 2% of the accessions were classified as resistant and 8% as moderately resistant (Phillips-Mora et al., 2009). However, follow-up studies at CATIE identified 50 diverse clones with moderate resistance to FPR and one resistant clone, the artificial hybrid CATIE R6 (UF-273 T1 × PA-169) (Phillips-Mora et al., 2017). In addition, field evaluations of UF 273 × Pa 169 progenies in Mexico showed a 41.27% reduction in FPR incidence compared to the average phenotypic value of progeny trials that included 22 full-sib families (Bonilla et al., 2021).

Cacao resistance to FPR should consider the pathogen variability observed in different locations. For example, a recent study of FRP diversity in tropical America (including 228 isolates obtained from cacao farms in 6 countries) used SSR and SNP markers to detect 6 hypothetical ancestral populations of *M. royeri* (Díaz-Valderrama et al., 2022). Two hotspots of *M. royeri* genetic diversity were observed in Colombian and Ecuadorian cacao plantations (Díaz-Valderrama et al., 2022). Thus, cacao clones should be tested for disease resistance in different countries while considering pathogen variability (Gutiérrez et al., 2016). In addition, preventive breeding in cultivated areas where FPR is not yet present, such as west Africa, is necessary (Gutiérrez et al., 2016).

Ecuador and Colombia are ideal countries for breeding resistance to FPR, considering the introduction of different strains of *M. royeri* in areas of intense cacao cultivation (Díaz-Valderrama et al., 2022; Espinoza-Lozano et al., 2022; Jaimes et al., 2016). For example, artificial inoculation of five cultivated clones with seven isolates of *M. royeri* in Colombia identified the hybrid ICS 95 (Amelonado × Criollo) as a resistant clone (Phillips-Mora et al., 2005). This clone and its derivatives (e.g., CCN 51) are a common source of resistance to FPR in Latin America (Bailey et al., 2018). In addition, agronomic evaluations of diverse clones in cultivated areas of Colombia identified several cacao clones with resistance to FPR (Cubillos, 2017; Jaimes et al., 2019; Osorio-Guarín et al., 2020). Evaluation of the breeding population representing CCN 51 × EET Nacional-type clones in coastal Ecuador identified four high yield clones resistant to FPR, which also exhibited resistance to WB and BPR (Jaimez et al., 2020). Nacional admixture genotypes showed resistance to FPR also in Ecuador (McElroy et al., 2018). Unless disease resistant clones are delivered to farmers, managing that FPR relies on extra management practices, such as removing infected pods from plantations (Cubillos, 2017; Phillips-Mora et al., 2012).

4.1.4 Genomic resources for breeding disease resistance

GWAS studies and QTL analysis in cacao have identified different loci associated with resistance to FPR, BPR, and WBD (Gutiérrez, Puig, et al., 2021; McElroy et al., 2018; Osorio-Guarín et al., 2020; Romero-Navarro et al., 2017). In Costa Rica at CATIE, six genes located on chromosomes 4, 5, 6, and 10 were associated with FPR resistance (Romero-Navarro et al., 2017). A higher proportion of Nacional ancestry within the 148 accessions analyzed was associated with resistance to FPR but susceptibility to BPR (Romero-Navarro et al., 2017). Follow-up analysis at CATIE using artificial inoculation in the F1 population from the Pound 7 × UF 273 cross identified six QTL associated with resistance to FPR (Gutiérrez, Puig, et al., 2021). Six SNPs located in chromosomes 2, 4, 7, 8, 9, and 10 accounted for 52.6% of

phenotypic variation (Gutiérrez, Puig, et al., 2021). The same study identified four QTL associated with resistance to BPR; four SNPs located on chromosomes 2, 4, 8, and 10 accounted for 50.6% of the phenotypic variation (Gutiérrez, Puig, et al., 2021).

In Ecuador, the natural occurrence of FPR and WB was observed among Nacional admixture genotypes, seven SNPs associated with FPR resistance located on chromosomes 1, 3, 5, 9, and 10, and nine SNPs associated with WBD resistance located on chromosomes 1, 2, 6, 7, and 9 were identified as markers for GS, which demonstrated a higher predictive accuracy of disease resistant plants when compared to single-marker selection (McElroy et al., 2018). For the cacao germplasm collection hosted in Colombia, two SNP markers were associated with FPR resistance on chromosomes 1 and 2, respectively, and four with WBD resistance on chromosomes 2 and 3 (Osorio-Guarín et al., 2020).

Although important economic diseases persist in cacao, wild populations exhibit disease resistance (Gutiérrez et al., 2016; Phillips-Mora et al., 2017). Only a few accessions have been used as a source of disease resistance to develop commercial clones, but more accessions should be studied for polygenic traits, such as disease resistance (Bailey et al., 2018; Gutiérrez et al., 2016). Meanwhile, using standard methods for artificial pod and leaf inoculation in germplasm collections and genotypes from different populations revealed new sources of disease resistance to FPR and BPR (Fister et al., 2020; Phillips-Mora et al., 2017; Puig et al., 2021). In addition, infection under field conditions also revealed new resistant genotypes. For example, the wild accession EBC-09 collected in the Colombian Amazon is resistant to FPR but not used commercially yet (Osorio-Guarín et al., 2020). Table 1 presents a summary of different candidate genes associated with disease resistance in cacao. Analysis of variation in these candidate genes in different genotypes can reveal potentially useful alleles for breeding programs. Thus, considering the continued demand for chocolate, breeding disease resistant varieties and renovating current cacao farms are more sustainable approaches than increasing

yield by planting new areas (Gutiérrez et al., 2016; Marelli et al., 2019; Phillips-Mora et al., 2012).

4.2 Genetic basis of quality traits

4.2.1 Chocolate flavor

A growing demand for chocolate also raises the need for related quality traits, including seed weight, fat content, low Cd content, and favorable flavor (CAOBISCO/ECA/FCC, 2015). The complex chocolate flavor depends not only on genotypes but also on soils, climate, farm management, bean fermentation, and roasting methods (Afoakwa et al., 2008; Kadow et al., 2013; Papalexandratou & Nielsen, 2016; Seguíne & Meinhardt, 2014). Cacao seeds are broadly classified for fine aroma that has economic implications for chocolate products and the international cacao market (Afoakwa et al., 2008), where cacao with better aroma is higher in price compared to standard cacao. Only 12% of the global production is considered having a fine flavor. Fine chocolate flavor is commonly attributed to two genetic groups, Criollo and Nacional, and the hybrid Trinitarios (Criollo × Amelonado) (ICCO, 2022).

In the case of Nacional, the mature fruits are green due to the absence of anthocyanins and have a fruity–floral taste. Contrary, mature red fruits produced by standard cultivars such as CNN 5 are associated with unfavorable flavor (Motamayor et al., 2013). Based on the genome sequence of cultivar Matina1-6 (Amelonado group), association, and gene expression analysis, it was possible to identify polymorphisms in the candidate gene TcMYB113 responsible for fruit color variation (Motamayor et al., 2013). Thus, a single SNP associated with the absence of anthocyanins in mature green fruits allows to select for favorable flavor in breeding populations (Motamayor et al., 2013).

Fine chocolate flavor (floral and fruity) in Nacional is genetically controlled by different candidate genes distributed along the 10 cacao chromosomes (Colonges et al., 2021; Colonges,

Jimenez, et al., 2022). Two principal biosynthesis pathways are associated with floral taste, monoterpene biosynthesis, and the L-phenylalanine degradation (Colonges et al., 2021). In total, 27 candidate genes were associated with floral aroma in cacao, which were identified by GWAS studies and sensory analysis of 152 modern Nacional cultivars of Ecuador (Colonges et al., 2021). Furthermore, three additional biosynthesis pathways are involved in the synthesis of the fruity aroma of cacao: pyrazine production, sugar, and fatty acid (FA) degradation (Colonges, Jimenez, et al., 2022). In total, 327 candidate genes were associated with the fruity aroma in cacao (Colonges, Jimenez, et al., 2022).

Follow-up studies observed 1824 candidate genes determining cacao aroma in wild accessions of Ecuadorian Amazon (Colonges, Loor Solorzano, et al., 2022). Among these genes, 53 were also identified in previous studies investigating aroma traits in cultivated Nacional (Colonges, Loor Solorzano, et al., 2022). In addition, 81 candidate genes were identified as determinants of astringency, nonvolatile compounds, and cacao bitterness (Colonges, Seguíne, et al., 2022). Overall, the balance among all these compounds provides the genetic basis for fine flavor formation in cacao and reveals a complex nature. Development of genetic markers based on candidate genes with high effects on favorable flavor would be necessary to specialized chocolate markets.

4.2.2 Cadmium (Cd) content

Cd is a heavy metal accumulated in edible parts of crops when transported from soils by the root system. The accumulation of Cd in food is neither a new problem nor unique to cacao. Fortunately, there is natural variation in Cd accumulation within crop species, and successful selection and breeding for low Cd-containing crops have been reported in barley, durum wheat, sunflower, rice, and soybean (Chen et al., 2007; Clemens et al., 2013; Grant et al., 2008; Lei et al., 2020).

The Cd transportation from soil to plant organs is associated with two main transporter genes encoding natural resistance-associated macrophage proteins and heavy metal ATPase (HMA), respectively (Clemens et al., 2013). For example, in rice (*Oryza sativa* L.), mutants with OsNRAMP5 defective function have successfully been used for breeding rice cultivars with low Cd accumulation and without affecting other agronomic traits (Ishikawa et al., 2012, 2019). Low Cd accumulation in rice was achieved by selecting varieties carrying functional alleles of OsHMA3, which promotes the Cd sequestration in root vacuoles, resulting in low Cd concentration in seeds (Ueno et al., 2010). Furthermore, an OsHMA3 allele with loss of function in Japonica rice cultivars was responsible for high Cd accumulation in grains (Yan et al., 2016). A transposable element presents in HvHMA3 allowed to enhance gene expression resulting in low Cd levels in barley grains (Lei et al., 2020). The low Cd allele originated from an international barley germplasm collection, being introduced into an elite cultivar by backcrossing, yielded grains with favorable agronomic traits, and low Cd accumulation (Lei et al., 2020). The introgressed barley lines yielded 0.1 mg Cd/kg compared to 0.4 mg Cd/kg in an elite cultivar; both groups of plants were grown in Cd contaminated soil (0.63 mg Cd/kg) (Lei et al., 2020).

Variation in cacao Cd accumulation called attention mainly because of international market restrictions for cacao powder and chocolate products with high Cd levels (between 0.10 and 0.80 mg Cd/kg) (The European Commission, 2014). The main factors related to the uptake of soil Cd by the cacao tree were soil pH, soil Cd availability, genotype, geographical location, and agronomic factors such as phosphate fertilizer (Oliveira et al., 2022). Different levels of Cd uptake have been observed among different cacao genotypes, suggesting genetic variation for seed Cd accumulation (Engbersen et al., 2019; Lewis et al., 2018). For example, in Trinidad and Tobago, 100 cacao accessions grown under the same conditions yielded different Cd contents in cacao seeds, ranging from 0.18 to 0.58 mg/kg for those classified as low Cd

accumulators, and from 1.72 to 2.16 mg/kg for high Cd accumulators (Lewis et al., 2018). Specifically, among the cacao genetic groups, the Contamana group exhibited high seed Cd content, whereas the Curaray group and hybrid populations such as Trinitario and Refractario are characterized by low Cd content (Lewis et al., 2018). Eleven cacao cultivars in Honduras demonstrated different Cd uptake rates (Engbersen et al., 2019). Accessions CAUCASIA-43, UF-676, and SPA-9 had high bean Cd accumulation (3–4 mg/kg), but UF 613 showed low bean Cd accumulation (~1.3 mg/kg) (Engbersen et al., 2019). In cacao farms of Ecuador, the main factor affecting Cd accumulation was the presence of high Cd in soils combined with the natural ability of Cd uptake observed in Nacional genotypes and CCN 51 cultivars (Argüello et al., 2019). Overall, there is a tendency of cultivated cacao to increase seed Cd content when trees are grown in soils with high Cd content, which is common in Latin America (Meter et al., 2019; Vanderschueren et al., 2021). This accumulation pattern occurs despite environmental factors, probably because of a limited frequency of low Cd alleles within cacao cultivars.

The genes associated with Cd uptake in cacao have been studied recently. Expression and functional analysis of *TcNRAMP5*, *TcHMA3*, and *TcHMA2* genes have demonstrated them as main Cd transporters in cacao (Moore et al., 2020; Ullah et al., 2018). To our knowledge, there is a gap between GWAS of Cd accumulation in seeds of cultivated cacao and genetic markers used for low Cd accumulation breeding. Cacao genotypes with low seed Cd accumulation (e.g., Red Amel 1/30, JA 5/39 and CL 27/50) can be tested as rootstocks for commercial clones to evaluate the genetic control of Cd accumulation (Lewis et al., 2018). However, soil–rootstock–scion interactions require further examination to validate potential low Cd cacao rootstocks (Fernández-Paz et al., 2021). For example, the commonly cultivated clone CCN 51 used as scion and the rootstock BN 34 accumulated higher Cd in root systems of juvenile cacao plants compared to seven different scion-rootstock combinations (Almeida et al., 2022).

4.2.3 Fat content and composition

The fat content of cacao seeds and their composition are vital for chocolate and cosmetics industries, and they demonstrated genetic and geographic variation (Almeida et al., 2008; Pires et al., 1998). Seed fat content varies among germplasm accessions from 45% to 60% in Bahia, Brazil (Pires et al., 1998), 51% to 58% in Rondônia, Brazil (Almeida et al., 2008), and 44% to 59% in Trinidad and Tobago (Khan et al., 2008). Cacao from different farms also varies for fat content with cacao cultivated in Ecuador showing a fat range of 45%–52%, whereas plantations in northeastern Peru show a low-fat content of 17%–30%. This variation was explained by geographic origin of samples and cultivated genotypes (Oliva-Cruz et al., 2021; Samaniego et al., 2021). On average, chocolate manufactures prefer a high fat content of 55%–58% in cacao seeds (CAOBISCO/ECA/FCC, 2015).

The fat profile of cacao is composed of three main FAs, which together account for more than 80% of the fat composition: saturated palmitic and stearic acids (C16:0, both with a 16-carbon chain, where the number before the colon specify the number of carbon atoms, and one after the colon, the number of double bonds) and monounsaturated palmitic oleic (C18:1). Other FAs are present in lower proportion, including saturated arachidic acid (C20:0) and the unsaturated linoleic acid (C18:2) (Mustiga et al., 2019; Samaniego et al., 2021; Torres-Moreno et al., 2015). This unique combination of FAs defines the texture, hardness, and an ideal melting point (37°C) of chocolate products at the mouthfeel (Afoakwa, 2016; CAOBISSCO/ECA/FCC, 2015). A high ratio of stearic acid increases the hardness of cacao butter (Vieira et al., 2015), whereas higher levels of oleic acid produce soft cacao butter being less appropriate for chocolate manufacture (CAOBISCO/ECA/FCC, 2015; Vieira et al., 2015). Cacao yields soft butter when it grows in cooler climates (~24°C) revealing a strong effect of temperature on cacao butter profile (Vieira et al., 2015).

The fat profile of cacao also defines the nutritional characteristics of related products and is under genetic and environmental control. A healthier fat profile includes lower amounts of saturated and higher amounts of unsaturated fat, being preferred in chocolate products (Torres-Moreno et al., 2015). Specifically, a lower amount of palmitic acid in cacao fat is ideal because it reduces the risk of cardiovascular health problems, and a higher amount of stearic acid is favored because it is desaturated in the organism as oleic acid (Oliva-Cruz et al., 2021; Samaniego et al., 2021). Variations in fat profiles are observed in cacao populations. For example, the amount of palmitic acid is low in Criollo (20%–30%) but slightly higher in Nacional (26%–29%) and in accessions with Iquitos ancestry (26%–34%) (Mustiga et al., 2019; Samaniego et al., 2021). However, cacao grown at higher temperature (~30–31°C) tends to produce a higher amount of palmitic acids, revealing a climate control (Mustiga et al., 2019). Furthermore, there is a negative correlation between total palmitic acid and the sum of stearic oleic and linoleic acids, which can be exploited to select genotypes with a healthier fat profile (Mustiga et al., 2019; Samaniego et al., 2021).

There are 84 genes orthologous to *Arabidopsis* that are potentially involved in fat production in cacao, of which 13 genes are unique to cacao (Argout et al., 2011). Specifically, the stearyl-acyl carrier protein desaturase (SAD) gene family regulates fat synthesis and its composition in cacao (Li et al., 2019; Mustiga et al., 2019; Zhang et al., 2015). The activity of TcSAD defines stable levels of unsaturated and saturated FAs in cacao (Zhang et al., 2015). Among the eight putative isoforms of TcSAD, the expression of TcSAD1 was positively correlated with oleic acid and fat synthesis (Zhang et al., 2015). TcSAD7 is also involved in the synthesis and accumulation of oleic acid in cacao (Li et al., 2019).

Based on a mapping population of 420 progenies derived from the TSH 1188 × CCN 51 cross analyzed with a 6K SNP array and fat profiles from 3292 seed samples, it was possible to identify a significant QTL for a higher level in palmitic acid. The QTL marker explained a

4.5% increase in palmitic acid, which average content was 29% within the fat profile (Mustiga et al., 2019). This mapping population showed a low heritability for fat content ($H^2 = 0.14$) mainly due to its complex nature and the effect of temperature and rainy seasons on seed fat production. However, it showed high heritability for palmitic acid ($H^2 = 0.43$) and stearic acid ($H^2 = 0.38$) (Mustiga et al., 2019), thus, demonstrating a good potential for breeding cacao genotypes with a healthier fat profile with lower palmitic and high stearic acids (Mustiga et al., 2019).

4.3 Genetic basis of agronomic traits

4.3.1 Seed traits and yield

The dry seeds of cacao are the valuable raw material for chocolate industries. Manufacturers require uniform and heavier seeds (at least 1.0 g per fermented dried seed) because lighter and smaller cacao seeds have less fat percentage and more shell (CAOBISCO/ECA/FCC, 2015). Thus, the average cacao seed weight is an important trait for breeding programs (Cilas et al., 2010; Doaré et al., 2020; Goenaga et al., 2009; Phillips-Mora et al., 2012). The seed weight varies among genotypes. It also depends on the number of seeds per fruit, seed position within the fruit, and fruiting cycle (Doaré et al., 2020). Fruits with a high number of seeds (≥ 40 per fruit) tend to have lighter seeds (< 0.94 g per seed) (Bekele et al., 2006). Seeds located in the apical sector of fruits also tend to be lighter, whereas longer fruiting cycles increase seed size and weight (Doaré et al., 2020).

Cacao populations show natural variation in seed weight. It ranged from 0.41 to 1.84 g per seed among 1900 accessions held in the International cacao collection of Trinidad (Bekele et al., 2020). Weight of wild cacao seeds ranged from 0.51 to 1.39 g compared to 0.55 to 1.88 g in cultivated clones (Bekele et al., 2006). In 11 Trinitario clones, even heavier seeds with a range of 1.15–1.84 g were identified (Bekele et al., 2020). In Brazil, seed weight followed a normal distribution across 221 clones and ranged from 0.3 to 2.8 g (Cilas et al., 2010). In

addition, it was possible to structure accessions based on ancestry. Those with Trinitario and Criollo ancestry (commonly cultivated) have heavier seeds (>1.70 g) as a response to artificial selection. Contrary, wild populations showed more variation in seed weight (e.g., 0.90–1.45 g in the Guiana population) (Cilas et al., 2010). The seed weight showed a high heritability ($H^2 = 0.50$), but the number of seeds per fruit had a lower heritability ($H^2 = 0.29$). Thus, it is possible to breed cacao trees with high seed weight (>1.5 g) rather than increased number of seeds per fruit (>38) (Cilas et al., 2010).

Cacao candidate genes determining seed and yield traits were identified on chromosome IV via QTL mapping (Fernandes et al., 2020). Nine out of 13 candidate genes were involved in specialized functions such as sugar transport; other genes were involved in carbohydrate, lipid, and glucose metabolism (Fernandes et al., 2020). For each agronomic trait, such as number of pods (fruits) harvested per tree, dry seed weight, average yield per tree (kg of dry seeds/year), and pod index, it was possible to link at least one significant SNP contributing to the phenotypic variation of these traits, respectively (Fernandes et al., 2020). Pod index refers to the number of fruits needed to produce 1 kg of dry seeds; a low pod index (e.g., 21–25) is preferred because it helps to select trees that yield heavier and larger seeds (Bekele et al., 2020; Goenaga et al., 2009). A QTL with the highest effect explained 23.6% of the variation in the number of healthy pods harvested. This trait also showed a higher heritability ($H^2 = 0.75$), thus, facilitating the selection of trees yielding more pods (Fernandes et al., 2020). Dry seed weight also exhibited a high heritability ($H^2 = 0.56$), and among 16 QTL observed, one explained 14% of the variation for this trait (Fernandes et al., 2020).

4.3.2 Self-incompatibility (SI)

Self-incompatibility (SI) prevents self-fertilization in plants and helps to avoid inbreeding depression and to maintain plant genetic variability. It is common among crops (e.g., coffee, sunflower, and potato) and tropical trees, including cacao (Finkeldey & Hattemer, 2007; López

et al., 2021). Among cacao populations, SI is the predominant phenotype, but Amelonado and Criollo genetic groups exhibit self-compatibility (SC), and it is a trait of economic interest because SC trees are more productive (Lanaud et al., 2017; López et al., 2021; Royaert et al., 2011). Although cacao produces a lot of flowers (~125.000 annually per tree), no more than 10% of them are pollinated and less than 5% develop into mature fruits (Falque et al., 1995; Groeneveld et al., 2010). Thus, wild and planted cacao rely on midges flies from the Ceratopogonidae family for natural cross pollination (da Silva et al., 2016; Young, 1994). SI in cacao occurs when gametes fail to fuse in the ovary, preventing fertilization followed by flower abscission (Cope, 1962). Lanaud et al. (2017) provided a detailed description of the cacao SI system. Briefly, the complex SI system in cacao depends on different candidate genes that act independently and located on chromosome 1 (Chr1) and chromosome 4 (Chr4). Loci on both chromosomes were associated with gametic selection (Lanaud et al., 2017). Chr1 acts prior to gamete fusion and includes three candidate genes. For two of them expression was observed in style tissues. Chr4 includes nine candidate genes identified with SI reactions, including fruit drop (Lanaud et al., 2017).

Because cacao SI limits yield, artificial (handmade) pollination within and among cultivars help to identify and interplant compatible trees in the field to improve fruit productivity (Akoa et al., 2021; Forbes et al., 2019; López et al., 2021). However, artificial pollination has drawbacks because it can be influenced by rain, wind, and insect attacks and requires mature blooming trees (da Silva et al., 2016). Another alternative is to select SC cacao seedlings at an early stage using genetic markers (da Silva et al., 2016; Lanaud et al., 2017; Royaert et al., 2011). For example, seven SNPs markers associated with SC were discovered based on GWAS analysis (da Silva et al., 2016). The study used 5301 SNPs distributed over 10 cacao chromosomes, 291 diverse cacao trees, and flower set as phenotype (da Silva et al., 2016). The seven SNPs were located in the proximal end of CH4, where at least two genes were associated

with pollen rupture (da Silva et al., 2016). In addition, Lanaud et al. (2017) developed two SSR markers (mSI_303 and mSI_460) located in CH4 for predicting SC in cacao breeding populations.

Overall, planting SC trees, whose selection can be assisted using genetic markers (e.g., SNPs or SSRs), will reduce extra cost of pollination labor, and commercial cacao farms will not depend on specific designs where compatible clones need to be interplanted.

5 GENOMIC RESOURCES AVAILABLE FOR CACAO BREEDING

A better understanding of the genes contributing to agronomic traits in cacao was obtained from the sequence and annotation of two reference genomes, Criollo B97-61 (Argout et al., 2011, 2017) and Matina1-6 (Motamayor et al., 2013). From annotation of these genomes, it was possible to identify key genes involved in quality traits (e.g., fat biosynthesis, fine flavor, and fruit color). Details of Criollo and Matina cacao genomes are presented by Figueira and Camargo-Scotton (2020). These genomes are highly homozygous and do not represent wild and high yield cultivated hybrids (Morrissey et al., 2019). Thus, Pound 7, a wild cacao tree widely used in breeding because its high yield and resistance to BPD was sequenced using the low-cost MinION sequencer (Morrissey et al., 2019). Applying this sequencing strategy to cultivars of high local adaptability, for example, CCN 51 (Boza et al., 2014; Jaimez et al., 2022), can guide breeding and gene editing in specific traits like reduced bitterness and undesirable flavor. Table 2 presents the genomic statistics of these three genomes.

TABLE 2. *Theobroma cacao* L. reference genomes and genomic statics.

Cacao genome	Estimated total genome size, Mbp	Total assembly, Mbp	Number of contigs	Number of scaffolds	Scaffold N50, Mbp	Number of genes	Reference
Criollo B97-61 V2	430	326	7743	431	36.4	28,798	Argout et al., 2017
Matina1-6	445	331	20,103	714	34.4	29,408	Motamayor et al., 2013
Pound 7 (minimap)	442	393	2720	506	37.5		Morrissey et al., 2019

Abbreviation: Mbp, million base pairs.

Advances in cacao biotechnology, such as tissue culture, genetic transformation, and gene editing, offer new tools that enable precise breeding in cacao (Figueira & Camargo-Scotton, 2020; Wickramasuriya & Dunwell, 2018). For example, a pioneer study proves that CRISPR can be used in cacao, and improved resistance to BPD was achieved by knockout of the TCNPR3 gene (Fister et al., 2018). New experiments are needed to test if CRISPR is effective to improve yield, considering the co-localization of candidate genes with three SNP markers significantly associated with pod index and seed size (Bekele et al., 2022). In addition, physiological traits for genetic editing include small plant architecture and the abscission of mature fruits (Figueira & Camargo-Scotton, 2020). Crown reduction and semi-dwarfism reduce pruning intervals (Pereira et al., 2017), and abscission of mature fruits facilitates harvesting (Kuhn et al., 2010). Both traits reduce management costs, but studies on the genetic basis and candidate genes for these two traits are missing.

The Infinium cacao chip containing 6K SNPs (Livingstone et al., 2015) and the upgraded 15K SNP array (Livingstone et al., 2017) have been used in cacao GWAS and QTL analysis for improving disease resistance (Fernandes et al., 2018; Gutiérrez, Puig, et al., 2021; Romero-Navarro et al., 2017; Royaert et al., 2016) and yield (Fernandes et al., 2020; Romero-Navarro et al., 2017). Another application of genome-wide SNP data includes GS for predicting disease resistance and yield in breeding populations (McElroy et al., 2018). In addition, 3412 SNPs, selected from the 6 and 15K chocolate SNP chip, were used to develop cacao inbred lines aiming to generate hybrids and exploit potential heterosis in cacao breeding (Lopes et al., 2022). Specifically, the genome-wide SNP data helped to select cacao clones with up to 90% homozygosity within a germplasm collection in Brazil, and three generations of selfing monitored with SNP data can increase homozygosity up to 95% in clones like CCN 51 (Lopes et al., 2022).

Standard SNPs for cacao germplasm management aim to improve the efficiency and diversity of accessions maintained in *ex situ* collections (Gutiérrez, Martínez, et al., 2021; Mahabir et al., 2020). For example, SNP genotyping was used for validating cacao germplasm identity and detection of off-types in breeding programs (DuVal et al., 2017; Olasupo et al., 2018; Wang et al., 2020).

6 CONCLUSIONS AND OUTLOOK

Cacao agroforestry provides long-term land use with increased forest cover that benefits rural economies in the tropics. In addition, planting genetically improved cacao reduces the pressure on new land to satisfy the international demand of chocolate. We propose here research areas on germplasm management and agronomic traits that will mostly contribute to breeding new varieties required for sustainable cacao cultivation.

6.1 Conservation and use of cacao genetic diversity

In this review, the discussed geographic distribution of wild cacao in Amazonia represents trees mostly observed and/or collected along main rivers, suggesting that large gaps exist in the distribution presented in available studies (Figure 1). Considering the complexity of the Amazon basin, there is room for further plant collections and search for new cacao populations and germplasm. Unfortunately, the progressive deforestation of Amazonia is eroding the natural habitat not only of cacao populations but also of many more native plant species with crop and economic potential (Coelho et al., 2021; Neves & Heckenberger, 2019; Schultes, 1979). Thus, more genetic studies are urgently needed to study new cacao populations using next generation high-throughput sequencing and genotyping. For instance, to improve efficiency of cacao conservation, a 219 SNP genotyping panel was recently developed (Gutiérrez, Martínez, et al., 2021) providing an efficient tool to study and manage the genetic diversity of wild and admixed accessions, including those that are maintained in around 47

cacao germplasm collections. Genetic gaps, off-types, and redundancy can be identified through SNP genotyping. Moreover, specific SNP profiles can identify commercial clones and parent trees carrying valuable agronomic traits in breeding programs.

Continued characterization of agronomic traits in different genetic groups (e.g., through genotyping accessions maintained in germplasm collections) can reveal new alleles valuable for breeding agronomic traits. It is important to note that multiyear agronomic characterization requires standard protocols shared among breeding programs for phenotyping traits, such as yield (e.g., pod index), seed weight, drought, and disease resistance. For example, new available methods can be adapted by different cacao research programs to evaluate yield and disease resistance (Jaimez et al., 2020), resistance to BPR in seedlings (Delgadillo-Durán et al., 2020), and on detached fruits (Puig et al., 2021). In addition, standard protocols to evaluate cacao quality and flavor are available (ISCQF, 2020). Overall, standard phenotyping is a priority that allows replicability and comparability of trait data among cacao breeding programs.

6.2 Disease resistance

Characterization of cacao resistance to FPR can benefit from the identified centers of genetic diversity of *M. roleri* in Magdalena valley of Colombia, Coastal Ecuador, and Ecuadorian Amazon. Thus, different cacao genetic groups can be tested against different *M. roleri* isolates, potentially identifying new sources of disease resistance. In addition, studying wild populations of *M. pernicioso* in Amazonia (e.g., Colombia, Peru, and Ecuador) can provide more information about genetic diversity, dispersals patterns, and genotypes of *M. pernicioso* causing WBD in cultivated cacao. Moreover, it will help to spot forest areas in Amazonia for collecting wild cacao that likely evolved resistance to WBD and BPR.

6.3 Climate adaptation

Breeding and planting cacao trees tolerant to drought and heat in West Africa can reduce the risk of deforestation associated with new suitable planting areas. Different cacao populations can be tested in controlled drought experiments aided by GWAS analysis to identify potential SNP markers in root phenotypes (e.g., angle and biomass) that improve water use efficiency. Of particular interest are the populations that originated along the border of Peru and Ecuador, and in the south of Peru, where wild trees grow naturally with precipitations around 1000 mm/year (Figure 2), potentially providing a natural source of germplasm to exploit drought tolerance traits.

6.3 Low seed cadmium (Cd) accumulation

Two main genes have been identified as main Cd transporters from soil to the cacao tree—TcNRAMP5 and TcHMA (Moore et al., 2020; Ullah et al., 2018). In addition, there is evidence of natural variation in seed Cd accumulation among cacao genotypes growing under the same soil conditions (Lewis et al., 2018). However, we neither know whether TcNRAMP5 and TcHMA gene variants are correlated with seed Cd accumulation levels, nor how these potential variants are distributed in cacao populations. By identifying gene variants (e.g., using SNP markers) associated with low and high seed Cd accumulation, breeders can develop new varieties to limit the Cd exposure on chocolate consumers.

6.5 Gene editing

Natural tree populations provide valuable genetic diversity for breeding programs. Using this variation and new genomic technologies, such as gene editing, it will be possible to accelerate tree genetic improvement (Cao et al., 2022). For example, the efficiency of gene editing using the CRISPR/Cas9 method was tested in cacao accession PSU SCA6 by knocking out the TcNPR3 gene that resulted in enhanced resistance to *Phytophthora tropicalis* (Fister et al.,

2018). Elite cacao clones can directly be improved by CRISPR-based gene editing, but it requires high quality reference genomes for optimizing specific gRNA constructs and to avoid off-target products (Morrissey et al., 2019). In addition, cacao tissue culture needs to be optimized, although protocols for somatic embryogenesis using staminode explants were adapted in nine cacao cultivars (Jones et al., 2022). In addition, further gene editing studies can validate candidate gene functions (e.g., loss of the function mutations) in important biological processes of cacao, such as disease resistance, and SC and Cd uptake. For example, functionally characterized Cd uptake causative genes (TcNRAMP5 and TcHMA) could be further validated in wild populations and clones. However, the potential of CRISPR-based technology is limited to agronomic traits controlled by a few major genes, genome editing of polygenic complex traits is a complicated task that may cause unintended results (Cao et al., 2022). Ultimately, the acceptance of gene editing and genetic modifications of cacao will depend on public acceptance and official regulations, but farmers and consumers will directly benefit from improved cacao.

6.6 Genomic selection (GS)

GS is more often implemented now in tree improvement due to the high accuracy of prediction of complex phenotypes that shortens breeding cycles (Grattapaglia, 2022). In cacao, GS using a 15K SNP genotyping array proved to be effective in improving quantitative traits, such as disease resistance and yield (Romero et al., 2017; McElroy et al., 2018). High-density SNP genotyping arrays can provide long-term benefits like higher genetic gain and better predictions (Grattapaglia, 2022). For example, medium to high-density SNP genotyping arrays are common in breeding tropical perennial crops, such as Eucalyptus (60K SNPs) and oil palm (200K SNPs) (Seyum et al., 2022). In addition, experimental approaches such as development of cacao inbred lines as elite parents can be used for mate selection and exploitation of hybrid vigor (Lopes et al., 2022), followed by GS of untested hybrids across different environments

(Grattapaglia, 2022). Methods and practical guidelines for designing GS programs specifically for cacao are much needed similar to other tree crops (see reviews by Grattapaglia, 2017, and Isik, 2022). The private sector and research agencies collaborated to test cacao hybrids and shorten the breeding cycle of disease resistance and yield under GS schemes (Romero et al., 2017; McElroy et al., 2018). However, challenges still remain, such as building a mature breeding population (e.g., by eliminating low yield and susceptible individuals at an early stage), high cost of genotyping, multiyear data collection and complexity of phenotyping disease resistance in training populations, and testing for $G \times E$ interactions in locations with higher disease pressure.

AUTHOR CONTRIBUTIONS

Helmuth E. Nieves-Orduña: Conceptualization; Funding acquisition; Investigation; Visualization; Writing — Original Draft; Writing — Review & Editing. Konstantin V. Krutovsky: Conceptualization; Investigation; Supervision; Validation; Writing — Review & Editing. Oliver Gailing: Conceptualization; Funding acquisition; Investigation; Project administration; Supervision; Validation; Writing — Review & Editing.

ACKNOWLEDGMENTS

We thank the German Academic Exchange Service-DAAD for the financial support provided to Helmuth E. Nieves-Orduña. We thank Prof. Dr. Hans ter Steege and the Amazon Tree Diversity Network for providing access to the observation points of *Theobroma cacao* L. in Amazonia. We acknowledge support by the Open Access Publication Funds of the Georg-August University of Göttingen. We thank anonymous reviewers for insightful suggestions, which helped us improve the manuscript, and editors for handling the manuscript.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

7 REFERENCES

- Abbott, P. C., Benjamin, T. J., Burniske, G. R., Croft, M. M., Fenton, M., Kelly, C. R., Lundy, M., Rodríguez Camayo, F., & Wilcox, M. D. (2018). *An analysis of the supply chain of cacao in Colombia*. Cali. Purdue University and the International Center for Tropical Agriculture (CIAT). <https://cgspace.cgiar.org/handle/10568/96636>
- Adu-Gyamfi, R., & Wetten, A. (2012). Cryopreservation of cocoa (*Theobroma cacao* L.) somatic embryos by vitrification. *CryoLetters*, 33(6), 494–505.
- Adu-Gyamfi, R., & Wetten, A. (2020). Cocoa (*Theobroma cacao* L.) somatic embryos tolerate some ice crystallization during cryopreservation. *Agricultural Sciences*, 11(03), 223–234. <https://doi.org/10.4236/as.2020.113014>
- Adu-Gyamfi, R., Wetten, A., & Marcelino Rodríguez López, C. (2016). Effect of cryopreservation and post-cryopreservation somatic embryogenesis on the epigenetic fidelity of cocoa (*Theobroma cacao* L.). *PLoS One*, 11(7), e0158857. <https://doi.org/10.1371/journal.pone.0158857>
- Afoakwa, E. O. (2016). *Chocolate science and technology* (2nd ed.). Wiley.
- Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and character in cocoa and chocolate: A critical review. *Critical Reviews in Food Science and Nutrition*, 48(9), 840–857. <https://doi.org/10.1080/10408390701719272>
- Akinagbe, A., Gailing, O., Finkeldey, R., & Lawal, A. (2019). Towards conservation of genetic variation of tropical tree species with differing successional status: The case of *Mansonia altissima* A. Chev and *Triplochiton scleroxylon* K. Schum. *Tropical Conservation Science*, 12, 1–9. <https://doi.org/10.1177/1940082919864267>
- Akoa, S. P., Onomo, P. E., Ndjaga, J. M., Ondobo, M. L., & Djocgoue, P.F. (2021). Impact of pollen genetic origin on compatibility, agronomic traits, and physicochemical quality of cocoa (*Theobroma cacao* L.) beans. *Scientia Horticulturae*, 287, 110278. <https://doi.org/10.1016/j.scienta.2021.110278>
- Ali, S. S., Shao, J., Strem, M. D., Phillips-Mora, W., Zhang, D., Meinhardt, L. W., & Bailey, B. A. (2015). Combination of RNAseq and SNP nanofluidic array reveals the center of genetic diversity of cacao pathogen *Moniliophthora roreri* in the upper Magdalena Valley of Colombia and its clonality. *Frontiers in Microbiology*, 6, 850. <https://doi.org/10.3389/fmicb.2015.00850>
- Allen, J. B. (1988). Geographical variation and population biology in wild *Theobroma cacao* [PhD]. The University of Edinburgh.
- Almeida, C. (1983). *Relatorio da expedicao botanica para coleta de cacau silvestre na regio de Ouro Preto d'Oeste, Rondonia*. CEPLAC/DEPEA. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=ALM83A
- Almeida, C. (2001). Ecology of natural populations. In L. Dias (Ed.), *Genetic improvement of cacao*. Food and Agriculture Organization of the United Nations. <https://ecoport.org/ep?SearchType=earticleView&earticleId=197&page=-2#section2690>
- Almeida, C., Barriga, J., Machado, P., & Bartley, B. (1987). *Evolucao do programa de conservacao dos recursos geneticos de cacau na Amazonia Brasileira*. Boletim Técnico No. 5. Ministério da Agricultura. Comissão Executiva do Plano da Lavoura Cacaueira.

- Almeida, C., Neto, F., Silva, P., Medeiros, J., & Lima, A. (2015). Genetic resources of cacao in Rondônia: Historical overview, origin and inventory. *Agrotropica*, 27(2), 93–124. https://www.ceplac.gov.br/paginas/agrotropica/revistas/agrotropica_27_2.pdf (In Portuguese with English Abstract)
- Almeida, C., Facundo, V. A., Brazil, H., Dias, L., Medeiros, J., & Costa, J. (2008). Diversity of cacao fat in germplasm from Rondônia, Brazil. *Crop Breeding and Applied Biotechnology*, 8(2), 141–148. <https://doi.org/10.12702/1984-7033.v08n02a07>
- Almeida, C., Machado, P., Barriga, J., & Silva, F. (1995). *Coleta da cacau (Theobroma cacao) da amazonia brasileira: uma abordagem historica e analitica*. CEPLAC, SRAO. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=ALM95A
- de Almeida, N., Almeida, A., Santos, N., Nascimento, J., de Carvalho, C., Pirovani, C., Ahnert, D., & Baligar, V. (2022). Scion-rootstock interaction and tolerance to cadmium toxicity in juvenile *Theobroma cacao* plants. *Scientia Horticulturae*, 300, 111086. <https://doi.org/10.1016/j.scienta.2022.111086>
- Arevalo-Gardini, E., Meinhardt, L. W., Zuñiga, L. C., Juan, A. G., Lambert, M., & Dapeng, Z. (2019). Genetic identity and origin of “Piura Porcelana”—A fine-flavored traditional variety of cacao (*Theobroma cacao*) from the Peruvian Amazon. *Tree Genetics & Genomes*, 15(1), 11. <https://doi.org/10.1007/s11295-019-1316-y>
- Argout, X., Salse, J., Aury, J. M., Guiltinan, M., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S., Abrouk, M., Murat, F., Fouet, O., Poulain, J., Ruiz, M., Roguet, Y., Rodier-Goud, M., Barbosa-Neto, J. F., Sabot, F., . . . Lanaud, C. (2011). The genome of *Theobroma cacao*. *Nature Genetics*, 43(2), 101–108. <https://doi.org/10.1038/ng.736>
- Argout, X., Martin, G., Droc, G., Fouet, O., Labadie, K., Rivals, E., Aury, J. M., & Lanaud, C. (2017). The cacao Criollo genome v2.0: An improved version of the genome for genetic and functional genomic studies. *BMC Genomics*, 18(1), 730. <https://doi.org/10.1186/s12864-017-4120-9>
- Argüello, D., Chavez, E., Laurysen, F., Vanderschueren, R., Smolders, E., & Montalvo, D. (2019). Soil properties and agronomic factors affecting cadmium concentrations in cacao beans: A nationwide survey in Ecuador. *The Science of the Total Environment*, 649, 120–127. <https://doi.org/10.1016/j.scitotenv.2018.08.292>
- Armenteras, D., Dávalos, L. M., Barreto, J. S., Miranda, A., Hernández-Moreno, A., Zamorano-Elgueta, C., González-Delgado, T. M., Meza-Elizalde, M. C., & Retana, J. (2021). Fire-induced loss of the world’s most biodiverse forests in Latin America. *Science Advances*, 7(33), eabd3357. <https://doi.org/10.1126/sciadv.abd3357>
- Artero, A. S., Silva, J. Q., Albuquerque, P. S. B., Bressan, E. A., Leal, G.A., Sebbenn, A. M., Griffith, G. W., & Figueira, A. (2017). Spatial genetic structure and dispersal of the cacao pathogen *Moniliophthora perniciosa* in the Brazilian Amazon. *Plant Pathology*, 66(6), 912–923. <https://doi.org/10.1111/ppa.12644>
- Bailey, B. A., Evans, H. C., Phillips-Mora, W., Ali, S. S., & Meinhardt, L. W. (2018). *Moniliophthora roreri*, causal agent of cacao frosty pod rot. *Molecular Plant Pathology*, 19(7), 1580–1594. <https://doi.org/10.1111/mpp.12648>

- Barrera-Ramírez, J., Prado, V., & Solheim, H. (2019). Life cycle assessment and socioeconomic evaluation of the illicit crop substitution policy in Colombia. *Journal of Industrial Ecology*, 23(5), 1237–1252. <https://doi.org/10.1111/jiec.12917>
- Barriga, J. (1982). *Programa da expedicao botanica para coleta de cacau nativo na Amazonia brasileira – 1982*. CEPLAC/DEPEA. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=BAR82B
- Bartley, B. G. D. (2005). *The genetic diversity of cacao and its utilization*. CABI Pub.
- Baruah, I. K., Ali, S. S., Shao, J., Lary, D., & Bailey, B. A. (2021). Changes in gene expression in leaves of cacao genotypes resistant and susceptible to *Phytophthora palmivora* infection. *Frontiers in Plant Science*, 12, 780805. <https://doi.org/10.3389/fpls.2021.780805>
- Bayer, C., & Kubitzki, K. (2003). Malvaceae. In K. Kubitzki & C. Bayer (Eds.), *Flowering plants dicotyledons* (pp. 225–311). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-07255-4_28
- Bekele, F., Bekele, I., Butler, D., & Bidaisee, G. (2006). Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm from the International Cocoa Genebank, Trinidad. *Genetic Resources and Crop Evolution*, 53(5), 933–948. <https://doi.org/10.1007/s10722-004-6692-x>
- Bekele, F., Bidaisee, G., Allegre, M., Argout, X., Fouet, O., Boccara, M., Saravanakumar, D., Bekele, I., & Lanaud, C. (2022). Genome-wide association studies and genomic selection assays made in a large sample of cacao (*Theobroma cacao* L.) germplasm reveal significant marker-trait associations and good predictive value for improving yield potential. *PLoS One*, 17(10), e0260907. <https://doi.org/10.1371/journal.pone.0260907>
- Bekele, F., Bidaisee, G., Singh, H., & Saravanakumar, D. (2020). Morphological characterisation and evaluation of cacao (*Theobroma cacao* L.) in Trinidad to facilitate utilisation of Trinitario cacao globally. *Genetic Resources and Crop Evolution*, 67(3), 621–643. <https://doi.org/10.1007/s10722-019-00793-7>
- Bekele, F., & Phillips-Mora, W. (2019). Cacao (*Theobroma cacao* L.) breeding. In J. M. Al-Khayri, S. M. Jain, & D. V. Johnson (Eds.), *Advances in plant breeding strategies: Industrial and food crops* (Vol., 46, pp. 409–487). Springer International Publishing. https://doi.org/10.1007/978-3-030-23265-8_12
- Bennett, A. B. (2003). Out of the Amazon: *Theobroma cacao* enters the genomic era. *Trends in Plant Science*, 8(12), 561–563. <https://doi.org/10.1016/j.tplants.2003.10.004>
- Bhagwat, S. A., Willis, K. J., Birks, H. J. B., & Whittaker, R. J. (2008). Agroforestry: A refuge for tropical biodiversity? *Trends in Ecology & Evolution*, 23(5), 261–267. <https://doi.org/10.1016/j.tree.2008.01.005>
- Blaser, W. J., Oppong, J., Hart, S. P., Landolt, J., Yeboah, E., & Six, J. (2018). Climate-smart sustainable agriculture in low-to-intermediate shade agroforests. *Nature Sustainability*, 1(5), 234–239. <https://doi.org/10.1038/s41893-018-0062-8>
- Bonilla, J. L. S., Lopes, U. V., Colmenero, A. Z., Valencia, B. B. M., Arrazate, C. H. A., & Gramacho, K. P. (2021). Estimation of genetic parameters associated with frosty pod rot (*Moniliophthora roreri*) and cacao production in Mexico. *Tree Genetics & Genomes*, 17(3), 24. <https://doi.org/10.1007/s11295-021-01498-8>

- Bowers, J. H., Bailey, B. A., Hebbar, P. K., Sanogo, S., & Lumsden, R. D. (2001). The impact of plant diseases on world chocolate production. *Plant Health Progress*, 2(1), 12. <https://doi.org/10.1094/PHP-2001-0709-01-RV>
- Boza, E. J., Motamayor, J. C., Amores, F. M., Cedeño-Amador, S., Tondo, C. L., Livingstone, D. S., Schnell, R. J., & Gutiérrez, O. A. (2014). Genetic characterization of the cacao cultivar CCN 51: Its impact and significance on global cacao improvement and production. *Journal of the American Society for Horticultural Science*, 139(2), 219–229. <https://doi.org/10.21273/JASHS.139.2.219>
- Bright, C., & Sarin, R. (2004). Forest cocoa could save endangered rainforest. *Appropriate Technology*, 31(1), 38–39.
- CacaoNet. (2012). *A global strategy for the conservation and use of cacao genetic resources, as the foundation for a sustainable cocoa economy*. Bioversity International.
- CAOBISCO/ECA/FCC. (2015). *Cocoa beans: Chocolate and cocoa industry quality requirements*. CAOBISCO/ECA/FCC. <https://cocoaquality.eu>
- Cao, H. X., Vu, G. T. H., & Gailing, O. (2022). From genome sequencing to CRISPR-based genome editing for climate-resilient forest trees. *International Journal of Molecular Sciences*, 23(2), 966. <https://doi.org/10.3390/ijms23020966>
- Cardoso, D., Särkinen, T., Alexander, S., Amorim, A. M., Bittrich, V., Celis, M., Daly, D. C., Fiaschi, P., Funk, V. A., Giacomini, L. L., Goldenberg, R., Heiden, G., Iganci, J., Kelloff, C. L., Knapp, S., Cavalcante de Lima, H., Machado, A. F. P., Dos Santos, R. M., Mello-Silva, R., Forzza, R. C. (2017). Amazon plant diversity revealed by a taxonomically verified species list. *Proceedings of the National Academy of Sciences of the United States of America*, 114(40), 10695–10700. <https://doi.org/10.1073/pnas.1706756114>
- Carletto, G. (1973). Expedicao internacional a Amazonia equatoriana para coleta de material botanico de cacau. *Theobroma*, 3, 41–47. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=CAR73A
- Castro-Nunez, A., Charry, A., Castro-Llanos, F., Sylvester, J., & Bax, V. (2020). Reducing deforestation through value chain interventions in countries emerging from conflict: The case of the Colombian cocoa sector. *Applied Geography*, 123, 102280. <https://doi.org/10.1016/j.apgeog.2020.102280>
- Cavers, S., Navarro, C., & Lowe, A. (2004). Targeting genetic resource conservation in widespread species: A case study of *Cedrela odorata* L. *Forest Ecology and Management*, 197(1–3), 285–294. <https://doi.org/10.1016/j.foreco.2004.05.019>
- Ceccarelli, V., Fremout, T., Zavaleta, D., Lastra, S., Imán Correa, S., Arévalo-Gardini, E., Rodriguez, C. A., Cruz Hilacondo, W., & Thomas, E. (2021). Climate change impact on cultivated and wild cacao in Peru and the search of climate change-tolerant genotypes. *Diversity and Distributions*, 27(8), 1462–1476. <https://doi.org/10.1111/ddi.13294>
- Chalmers, W. (1968). *Cacao collecting trip to the Oriente of Ecuador*. Trinidad. Cocoa Research Unit, University of the West Indies. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=CHA68A
- Chalmers, W. (1969). *Report of a cocoa collecting trip to the Oriente of Ecuador (May 20th–26th 1969)*. Trinidad. Cocoa Research Unit, University of the West Indies. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=CHA69A

- Chalmers, W. (1972). *INIAP-UWI expedition to collect Theobroma cacao in the Oriente region of Ecuador (June 25th–August 10th 1972)*. Trinidad. Cocoa Research Unit, University of the West Indies. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=CHA72B
- Chalmers, W. (1973). *Report on a visit to Ecuador, Peru, and Brazil (March 1st–April 15th 1973)*. Trinidad. Cocoa Research Unit, University of the West Indies. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=CHA73A
- Cheesman, E. E. (1944). Notes on the nomenclature, classification and possible relationships of cacao populations. *Tropical Agriculture*, 21(8), 144–159. <https://journals.sta.uwi.edu/ta/index.asp?action=viewIssue&issueId=435>
- Chen, F., Dong, J., Wang, F., Wu, F., Zhang, G., Li, G., Chen, Z., Chen, J., & Wei, K. (2007). Identification of barley genotypes with low grain Cd accumulation and its interaction with four microelements. *Chemosphere*, 67(10), 2082–2088. <https://doi.org/10.1016/j.chemosphere.2006.10.014>
- Chia Wong, J. A., Clement, D. P. L., Mournet, P., dos Santos Nascimento, A., Solís Bonilla, J. L., Lopes, U. V., Pires, J. L., & Gramacho, K. P. (2022). A high-density genetic map from a cacao F2 progeny and QTL detection for resistance to witches' broom disease. *Tree Genetics & Genomes*, 18(4), 31. <https://doi.org/10.1007/s11295-022-01562-x>
- Chumacero de Schawe, C., Durka, W., Tschardtke, T., Hensen, I., & Kessler, M. (2013). Gene flow and genetic diversity in cultivated and wild cacao (*Theobroma cacao*) in Bolivia. *American Journal of Botany*, 100(11), 2271–2279. <https://doi.org/10.3732/ajb.1300025>
- Cilas, C., Machado, R., & Motamayor, J.-C. (2010). Relations between several traits linked to sexual plant reproduction in *Theobroma cacao* L.: Number of ovules per ovary, number of seeds per pod, and seed weight. *Tree Genetics & Genomes*, 6(2), 219–226. <https://doi.org/10.1007/s11295-009-0242-9>
- Clemens, S., Aarts, M. G. M., Thomine, S., & Verbruggen, N. (2013). Plant science: The key to preventing slow cadmium poisoning. *Trends in Plant Science*, 18(2), 92–99. <https://doi.org/10.1016/j.tplants.2012.08.003>
- Clement, C. R. (1989). A center of crop genetic diversity in western Amazonia. *BioScience*, 39(9), 624–631. <https://doi.org/10.2307/1311092>
- Clement, C. R. (1999). 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. *Economic Botany*, 53(2), 188. <https://doi.org/10.1007/BF02866498>
- Clement, D. (1986). Cacaoyers de Guyane. Prospections. *Café, Cacao, Thé*, 30(1), 11–36. <https://agritrop.cirad.fr/462886>
- Cocoa Research Unit. (1995). *Report for 1995*. Trinidad. The University of the West Indies. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=CRU95A
- Coe, S. D., & Coe, M. D. (2013). *The true history of chocolate* (3rd ed.). Thames & Hudson.
- Coelho, S. D., Levis, C., Baccaro, F. B., Figueiredo, F. O. G., Pinassi Antunes, A., Steege, H., Peña-Claros, M., Clement, C. R., & Schiatti, J. (2021). Eighty-four per cent of all Amazonian arboreal plant individuals are useful to humans. *PLoS One*, 16(10), e0257875. <https://doi.org/10.1371/journal.pone.0257875>

- Colonges, K., Jimenez, J.-C., Saltos, A., Seguine, E., Loor Solorzano, R. G., Fouet, O., Argout, X., Assemat, S., Davrieux, F., Cros, E., Boulanger, R., & Lanaud, C. (2021). Two main biosynthesis pathways involved in the synthesis of the floral aroma of the Nacional cocoa variety. *Frontiers in Plant Science*, *12*, 681979. <https://doi.org/10.3389/fpls.2021.681979>
- Colonges, K., Jimenez, J.-C., Saltos, A., Seguine, E., Loor Solorzano, R., Fouet, O., Argout, X., Assemat, S., Davrieux, F., Cros, E., Lanaud, C., & Boulanger, R. (2022). Integration of GWAS, metabolomics, and sensorial analyses to reveal novel metabolic pathways involved in cocoa fruity aroma GWAS of fruity aroma in *Theobroma cacao*. *Plant Physiology and Biochemistry*, *171*, 213–225. <https://doi.org/10.1016/j.plaphy.2021.11.006>
- Colonges, K., Loor Solorzano, R. G., Jimenez, J.-C., Lahon, M.-C., Seguine, E., Calderon, D., Subia, C., Sotomayor, I., Fernández, F., Lebrun, M., Fouet, O., Rhoné, B., Argout, X., Costet, P., Lanaud, C., & Boulanger, R. (2022). Variability and genetic determinants of cocoa aromas in trees native to South Ecuadorian Amazonia. *Plants, People, Planet*, *4*(6), 618–637. <https://doi.org/10.1002/ppp3.10268>
- Colonges, K., Seguine, E., Saltos, A., Davrieux, F., Minier, J., Jimenez, J.-C., Lahon, M.-C., Calderon, D., Subia, C., Sotomayor, I., Fernández, F., Fouet, O., Rhoné, B., Argout, X., Lebrun, M., Costet, P., Lanaud, C., Boulanger, R., & Loor Solorzano, R. G. (2022). Diversity and determinants of bitterness, astringency, and fat content in cultivated Nacional and native Amazonian cocoa accessions from Ecuador. *The Plant Genome*, *15*, e20218. <https://doi.org/10.1002/tpg2.20218>
- Cope, F. W. (1962). The mechanism of pollen incompatibility in *Theobroma cacao* L. *Heredity*, *17*(2), 157–182. <https://doi.org/10.1038/hdy.1962.14>
- Coral, F. (1988). *Expedicao ao Vale do Rio Ucayali (1987–1988): Field notes and report of cacao germplasm collections from the Rio Ucayali drainage system in Peru*. http://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=COR88A
- Cornejo, O. E., Yee, M.-C., Dominguez, V., Andrews, M., Sockell, A., Strandberg, E., Livingstone, D., Stack, C., Romero, A., Umaharan, P., Royaert, S., Tawari, N. R., Ng, P., Gutierrez, O., Phillips, W., Mockaitis, K., Bustamante, C. D., & Motamayor, J. C. (2018). Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process. *Communications Biology*, *1*, 167. <https://doi.org/10.1038/s42003-018-0168-6>
- Cuatrecasas, J. (1964). Cacao and its allies: A taxonomic revision of the genus *Theobroma*. *Systematic Plant Studies*, *35*(6), 379–614. <https://repository.si.edu/handle/10088/27110>
- Cubillos, G. (2017). Frosty pod rot, disease that affects the cocoa (*Theobroma cacao*) crops in Colombia. *Crop Protection*, *96*, 77–82. <https://doi.org/10.1016/j.cropro.2017.01.011>
- da Silva, M. R., Clément, D., Gramacho, K. P., Monteiro, W. R., Argout, X., Lanaud, C., & Lopes, U. (2016). Genome-wide association mapping of sexual incompatibility genes in cacao (*Theobroma cacao* L.). *Tree Genetics & Genomes*, *12*(3), 15. <https://doi.org/10.1007/s11295-016-1012-0>
- Dand, R. (2011). *The international cocoa trade* (3rd ed.). Elsevier. <https://doi.org/10.1016/C2013-0-16163-6>
- Daymond, A. (2018). Safe handling and movement of cocoa germplasm for breeding. In P. Umaharan (Ed.), *Achieving sustainable cultivation of cocoa* (pp. 101–110). Burleigh Dodds Science Publishing. <https://doi.org/10.19103/AS.2017.0021.05>

- Daymond, A., & Bekele, F. (2022). Cacao. In P. M. Priyadarshan & S.M. Jain (Eds.), *Cash crops: Genetic diversity, erosion, conservation and utilization* (pp. 23–53). Springer International Publishing. https://doi.org/10.1007/978-3-030-74926-2_2
- Delgadillo-Durán, P., Soto-Suárez, M., Rodríguez-Polanco, L., Carrero-Gutierrez, M., Torres-Rojas, E., & Yockteng, R. (2020). A new method for the inoculation of *Phytophthora palmivora* (Butler) into cacao seedlings under greenhouse conditions. *Plant Methods*, *16*, 114. <https://doi.org/10.1186/s13007-020-00656-8>
- Desrosiers, R., & Buchwald, A. V. (1961). *Report of a trip to the Napo river*. Estación Experimental Tropical Pichilingue. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=DES61A
- Dias, L. A. S., Barriga, J. P., Kageyama, P. Y., & Almeida, C. (2003). Variation and its distribution in wild cacao populations from the Brazilian Amazon. *Brazilian Archives of Biology and Technology*, *46*(4), 507–514. <https://doi.org/10.1590/S1516-89132003000400003>
- Dias, L. A. S. (Ed.). (2001). *Genetic improvement of cacao*. Food and Agriculture Organization of the United Nations.
- Díaz-Valderrama, J. R., Leiva-Espinoza, S. T., & Aime, M. C. (2020). The history of cacao and its diseases in the Americas. *Phytopathology*, *110*(10), 1604–1619. <https://doi.org/10.1094/PHYTO-05-20-0178-RVW>
- Díaz-Valderrama, J. R., Zambrano, R., Cedeño-Amador, S., Córdova Bermejo, U., Casas, G. G., García-Zurita, N., Sánchez-Arévalo, J. A. J., Arévalo-Gardini, E., Dávila, D., Ruiz, J., Pinchi-Dávila, X., Quispe-Chacón, Z. R., Chia-Wong, J. A., Hurtado-Gonzales, O. P., Rodríguez-Callañaupa, C. A., Maldonado-Fuentes, C., Pérez-Callizaya, E., Leiva-Espinoza, S., Oliva-Cruz, M., Aime, M. C. (2022). Diversity in the invasive cacao pathogen *Moniliophthora roreri* is shaped by agriculture. *Plant Pathology*, *71*, 1721–1734. <https://doi.org/10.1111/ppa.13603>
- Doaré, F., Ribeyre, F., & Cilas, C. (2020). Genetic and environmental links between traits of cocoa beans and pods clarify the phenotyping processes to be implemented. *Scientific Reports*, *10*(1), 9888. <https://doi.org/10.1038/s41598-020-66969-9>
- Draper, F. C., Asner, G. P., Honorio Coronado, E. N., Baker, T. R., García-Villacorta, R., Pitman, N. C. A., Fine, P. V. A., Phillips, O. L., Zárate Gómez, R., Amasifuén Guerra, C. A., Flores Arévalo, M., Vásquez Martínez, R., Brienen, R. J. W., Monteagudo-Mendoza, A., Torres Montenegro, L. A., Valderrama Sandoval, E., Roucoux, K. H., Ramírez Arévalo, F. R., Mesones Acuy, Í., Baraloto, C. (2019). Dominant tree species drive beta diversity patterns in western Amazonia. *Ecology*, *100*(4), e02636. <https://doi.org/10.1002/ecy.2636>
- DuVal, A., Gezan, S. A., Mustiga, G., Stack, C., Marelli, J.-P., Chaparro, J., Livingstone, D., Royaert, S., & Motamayor, J. C. (2017). Genetic parameters and the impact of off-types for *Theobroma cacao* L. in a breeding program in Brazil. *Frontiers in Plant Science*, *8*, 2059. <https://doi.org/10.3389/fpls.2017.02059>
- Engbersen, N., Gramlich, A., Lopez, M., Schwarz, G., Hattendorf, B., Gutierrez, O., & Schulin, R. (2019). Cadmium accumulation and allocation in different cacao cultivars. *The Science of the Total Environment*, *678*, 660–670. <https://doi.org/10.1016/j.scitotenv.2019.05.001>

Espinoza-Lozano, F., Amaya-Márquez, D., Pinto, C. M., Villavicencio- Vásquez, M., Del Sosa Castillo, D., & Pérez-Martínez, S. (2022). Multiple introductions of *Moniliophthora roreri* from the Amazon to the pacific region in Ecuador and shared high azoxystrobin sensitivity. *Agronomy*, *12*(5), 1119. <https://doi.org/10.3390/agronomy12051119> The European Commission. (2014). Commission Regulation (EU) No 488/2014 of 12 May 2014 amending Regulation (EC) No 1881/2006

as regards maximum levels of cadmium in foodstuffs. *Official Journal of the European Union*, *138*, 75–79. <https://eur-lex.europa.eu/eli/reg/2014/488/oj>

Falque, M., Vincent, A., Vaissiere, B., & Eskes, A. (1995). Effect of pollination intensity on fruit and seed set in cacao (*Theobroma cacao* L.). *Sexual Plant Reproduction*, *8*(6), 354–360. <https://doi.org/10.1007/BF00243203>

Fang, J.-Y., Wetten, A., & Hadley, P. (2004). Cryopreservation of cocoa (*Theobroma cacao* L.) somatic embryos for long-term germplasm storage. *Plant Science*, *166*(3), 669–675. <https://doi.org/10.1016/j.plantsci.2003.11.002>

FAOSTAT. (2020). *Crops and livestock products: Cocoa, beans*. FAO-STAT. <https://www.fao.org/faostat/en/#data/QCL>

Fernandes, L. D. S., Correa, F. M., Ingram, K. T., Almeida, A. A., & Royaert, S. (2020). QTL mapping and identification of SNP- haplotypes affecting yield components of *Theobroma cacao* L. *Horticulture Research*, *7*, 26. <https://doi.org/10.1038/s41438-020-0250-3>

Fernandes, L. D. S., Royaert, S., Correa, F. M., Mustiga, G. M., Marelli, J.-P., Corrêa, R. X., & Motamayor, J. C. (2018). Mapping of a major QTL for *Ceratocystis Wilt* disease in an F1 population of *Theobroma cacao*. *Frontiers in Plant Science*, *9*, 155. <https://doi.org/10.3389/fpls.2018.00155>

Fernández-Paz, J., Cortés, A. J., Hernández-Varela, C. A., Mejía- de-Tafur, M. S., Rodríguez-Medina, C., & Baligar, V. C. (2021). Rootstock-mediated genetic variance in cadmium uptake by juvenile cacao (*Theobroma cacao* L.) genotypes, and its effect on growth and physiology. *Frontiers in Plant Science*, *12*, 777842. <https://doi.org/10.3389/fpls.2021.777842>

Figueira, A., & Camargo-Scotton, D. (2020). *Theobroma cacao* Cacao. In R. E. Litz, F. Pliego-Alfaro, & J. I. Hormaza (Eds.), *Biotechnology of fruit and nut crops* (pp. 282–313). CABI.

Finkeldey, R., & Hattemer, H. H. (2007). Tropical forest genetics. *Tropical forestry*. Springer.

Fister, A. S., Landherr, L., Maximova, S. N., & Gultinan, M. J. (2018). Transient expression of CRISPR/Cas9 machinery targeting *TcNPR3* enhances defense response in *Theobroma cacao*. *Frontiers in Plant Science*, *9*, 268. <https://doi.org/10.3389/fpls.2018.00268>

Fister, A. S., Leandro-Muñoz, M. E., Zhang, D., Marden, J. H., Tiffin, P., dePamphilis, C., Maximova, S., & Gultinan, M. J. (2020). Widely distributed variation in tolerance to *Phytophthora palmivora* in four genetic groups of cacao. *Tree Genetics & Genomes*, *16*(1), 1–9. <https://doi.org/10.1007/s11295-019-1396-8>

Fister, A. S., Mejia, L. C., Zhang, Y., Herre, E. A., Maximova, S. N., & Gultinan, M. J. (2016). *Theobroma cacao* L. pathogenesis related gene tandem array members show diverse expression dynamics in response to pathogen colonization. *BMC Genomics*, *17*, 363. <https://doi.org/10.1186/s12864-016-2693-3>

- Forbes, S. J., Mustiga, G., Romero, A., Northfield, T. D., Lambert, S., & Motamayor, J. C. (2019). Supplemental and synchronized pollination may increase yield in Cacao. *HortScience*, *54*(10), 1718–1727. <https://doi.org/10.21273/HORTSCI12852-18>
- Fouet, O., Loor Solorzano, R. G., Rhoné, B., Subía, C., Calderón, D., Fernández, F., Sotomayor, I., Rivallan, R., Colonges, K., Vignes, H., Angamarca, F., Yaguana, B., Costet, P., Argout, X., & Lanaud, C. (2022). Collection of native *Theobroma cacao* L. accessions from the Ecuadorian Amazon highlights a hotspot of cocoa diversity. *Plants, People, Planet*, *4*(6), 605–617. <https://doi.org/10.1002/ppp3.10282>
- Garcia-Davila, C., Aldana Gomero, D., Renno, J.-F., Diaz Soria, R., Hidalgo Pizango, G., Flores Llampazo, G., Castro-Ruiz, D., Mejia de Loayza, E., Angulo Chavez, C., Mader, M., Tysklind, N., Paredes- Villanueva, K., Del Castillo Torres, D., Degen, B., & Honorio Coronado, E. N. (2020). Molecular evidence for three genetic species of *Dipteryx* in the Peruvian Amazon. *Genetica*, *148*(1), 1–11. <https://doi.org/10.1007/s10709-019-00082-2>
- Gasic, K., Preece, J. E., & Karp, D. (2016). Register of new fruit and nut cultivars list 48. *HortScience*, *51*(6), 620–652. <https://doi.org/10.21273/HORTSCI.51.6.620>
- Gentry, A. H. (1982). Neotropical floristic diversity: Phytogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny? *Annals of the Missouri Botanical Garden*, *69*(3), 557. <https://doi.org/10.2307/2399084>
- Gentry, A. H. (1989). Speciation in tropical forests. In L. B. Holm- Nielsen, I. C. Nielsen, & H. Balslev (Eds.), *Tropical forests: Botanical dynamics, speciation and diversity: Symposium Papers* (pp. 113– 134). Elsevier. <https://doi.org/10.1016/B978-0-12-353550-4.50017-4>
- Goenaga, R., Irizarry, H., & Irish, B. (2009). TARS series of cacao germplasm selections. *HortScience*, *44*(3), 826–827. <https://doi.org/10.21273/HORTSCI.44.3.826>
- Gomes, V. H. F., Vieira, I. C. G., Salomão, R. P., & Steege, H. T. (2019). Amazonian tree species threatened by deforestation and climate change. *Nature Climate Change*, *9*(7), 547–553. <https://doi.org/10.1038/s41558-019-0500-2>
- González-Orozco, C. E., Galán, A. A. S., Ramos, P. E., & Yockteng, R. (2020). Exploring the diversity and distribution of crop wild relatives of cacao (*Theobroma cacao* L.) in Colombia. *Genetic Resources and Crop Evolution*, *67*, 2071–2085. <https://doi.org/10.1007/s10722-020-00960-1>
- Grant, C. A., Clarke, J. M., Duguid, S., & Chaney, R. L. (2008). Selection and breeding of plant cultivars to minimize cadmium accumulation. *The Science of the Total Environment*, *390*(2–3), 301–310. <https://doi.org/10.1016/j.scitotenv.2007.10.038>
- Grattapaglia, D. (2017). Status and perspectives of genomic selection in forest tree breeding. In R. K. Varshney, M. Roorkiwal, & M. E. Sorrells (Eds.), *Genomic selection for crop improvement* (pp. 199–249). Springer International Publishing. https://doi.org/10.1007/978-3-319-63170-7_9
- Grattapaglia, D. (2022). Twelve years into genomic selection in forest trees: Climbing the slope of enlightenment of marker assisted tree breeding. *Forests*, *13*(10), 1554. <https://doi.org/10.3390/f13101554>
- Groeneveld, J. H., Tschardtke, T., Moser, G., & Clough, Y. (2010). Experimental evidence for stronger cacao yield limitation by pollination than by plant resources. *Perspectives in*

Plant Ecology, Evolution and Systematics, 12(3), 183–191. <https://doi.org/10.1016/j.ppees.2010.02.005>

Gutiérrez, O., Campbell, A., & Phillips-Mora, W. (2016). Breeding for disease resistance in cacao. In B. A. Bailey & L. W. Meinhardt (Eds.), *Cacao diseases* (pp. 567–609). Springer International Publishing. https://doi.org/10.1007/978-3-319-24789-2_18

Gutiérrez, O., Martínez, K., Zhang, D., Livingstone, D. S., Turnbull, C. J., & Motamayor, J. C. (2021). Selecting SNP markers reflecting population origin for cacao (*Theobroma cacao* L.) germplasm identification. *Beverage Plant Research*, 1(1), 1–9. <https://doi.org/10.48130/BPR-2021-0015>

Gutiérrez, O., Puig, A., Phillips-Mora, W., Bailey, B. A., Ali, S. S., Mockaitis, K., Schnell, R. J., Livingstone, D., Mustiga, G., Royaert, S., & Motamayor, J. C. (2021). SNP markers associated with resistance to frosty pod and black pod rot diseases in an F1 population of *Theobroma cacao* L. *Tree Genetics & Genomes*, 17(3), 106. <https://doi.org/10.1007/s11295-021-01507-w>

Haffer, J., & Prance, G. (2001). Climatic forcing of evolution in Amazonia during the Cenozoic: On the refuge theory of biotic differentiation. *Amazoniana: Limnologia Et Oecologia Regionalis Systematis Fluminis Amazonas*, 16(3/4), 579–607.

Heming, N. M., Schroth, G., Talora, D. C., & Faria, D. (2022). Cabruca agroforestry systems reduce vulnerability of cacao plantations to climate change in southern Bahia. *Agronomy for Sustainable Development*, 42(3), 48. <https://doi.org/10.1007/s13593-022-00780-w>

Hernández-Núñez, H. E., Gutiérrez-Montes, I., Bernal-Núñez, A. P., Gutiérrez-García, G. A., Suárez, J. C., Casanoves, F., & Flora, C. B. (2022). Cacao cultivation as a livelihood strategy: Contributions to the well-being of Colombian rural households. *Agriculture and Human Values*, 39(1), 201–216. <https://doi.org/10.1007/s10460-021-10240-y>

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978. <https://doi.org/10.1002/joc.1276>

Hoang, N. T., & Kanemoto, K. (2021). Mapping the deforestation footprint of nations reveals growing threat to tropical forests. *Nature Ecology & Evolution*, 5(6), 845–853. <https://doi.org/10.1038/s41559-021-01417-z>

Hor, Y. L., Chin, H. F., & Karim, M. (1984). The effect of seed moisture and storage temperature on the storability of cocoa (*Theobroma cacao*) seeds. *Seed Science and Technology*, 12(2), 415–420.

Hunter, J. R. (1990). The status of cacao (*Theobroma cacao*, Sterculiaceae) in the western hemisphere. *Economic Botany*, 44(4), 425–439. <https://doi.org/10.1007/BF02859775>

ICCO. (2021). *Production of cocoa beans: Quarterly bulletin of cocoa statistics, Vol. XLVII, No. 4, Cocoa year 2020/21*. The International Cocoa Organization (ICCO).

ICCO. (2022). *Fine or flavour cocoa*. ICCO. <https://www.icco.org/fine-or-flavor-cocoa/>

ISCQF. (2020). *First draft of the protocol for cocoa liquor sensory evaluation: Part of the International Standards for the Assessment of Cocoa Quality and Flavour (ISCQF)*. Compiled by Bioversity International, in Collaboration with the Members of the ISCQF Working Group. <https://www.cocoaqualitystandards.org/protocols-for-review/download>

- Isik, F. (2022). Genomic prediction of complex traits in perennial plants: A case for forest trees. In N. Ahmadi & J. Bartholome (Eds.), *Genomic prediction of complex traits: Methods and Protocols* (pp. 493–520). Springer Protocols: Methods in Molecular Biology, Volume 2467. Humana Press. https://doi.org/10.1007/978-1-0716-2205-6_18
- Ishikawa, S., Abe, T., Kuramata, M., & Hayashi, S. (2019). Development of low-cadmium accumulating rice. In S. Himeno & K. Aoshima (Eds.), *Cadmium toxicity. Current topics in environmental health and preventive medicine* (pp. 139–150). Springer. https://doi.org/10.1007/978-981-13-3630-0_11
- Ishikawa, S., Ishimaru, Y., Igura, M., Kuramata, M., Abe, T., Senoura, T., Hase, Y., Arao, T., Nishizawa, N. K., & Nakanishi, H. (2012). Ion- beam irradiation, gene identification, and marker-assisted breeding in the development of low-cadmium rice. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(47), 19166–19171. <https://doi.org/10.1073/pnas.1211132109>
- Jaimes, Y., Gonzalez, C., Rojas, J., Cornejo, O. E., Mideros, M. F., Restrepo, S., Cilas, C., & Furtado, E. L. (2016). Geographic differentiation and population genetic structure of *Moniliophthora roreri* in the principal cocoa production areas in Colombia. *Plant Disease*, *100*(8), 1548–1558. <https://doi.org/10.1094/PDIS-12-15-1498-RE>
- Jaimes, Y., Ribeyre, F., Gonzalez, C., Rojas, J., Furtado, E., & Cilas, C. (2019). Factors affecting the dynamics of frosty pod rot in the main cocoa areas of Santander State, Colombia. *Plant Disease*, *103*(7), 1665–1673. <https://doi.org/10.1094/PDIS-10-18-1761-RE>
- Jaimes, Y., Gonzalez, C., Rojas, J., Rivera, J. J., Cilas, C., & Argout, X. (2022). Population structure of *Moniliophthora perniciosa* in the main cacao producing departments of Colombia. *Plant Disease*, *106*(5), 1492–1501. <https://doi.org/10.1094/PDIS-12-20-2679-RE>
- Jaimez, R., Vera, D., Mora, A., Loor, R., & Bailey, B. (2020). A disease and production index (DPI) for selection of cacao (*Theobroma cacao*) clones highly productive and tolerant to pod rot diseases. *Plant Pathology*, *69*(4), 698–712. <https://doi.org/10.1111/ppa.13156>
- Jaimez, R. E., Barragan, L., Fernández-Niño, M., Wessjohann, L. A., Cedeño-García, G., Sotomayor Cantos, I., & Arteaga, F. (2022). *Theobroma cacao* L. cultivar CCN 51: A comprehensive review on origin, genetics, sensory properties, production dynamics, and physiological aspects. *PeerJ*, *10*, e12676. <https://doi.org/10.7717/peerj.12676>
- Jenderek, M. M., & Reed, B. M. (2017). Cryopreserved storage of clonal germplasm in the USDA National Plant Germplasm System. *In Vitro Cellular & Developmental Biology – Plant*, *53*(4), 299–308. <https://doi.org/10.1007/s11627-017-9828-3>
- Jezeer, R. E., Verweij, P. A., Santos, M. J., & Boot, R. G. (2017). Shaded coffee and cocoa – Double dividend for biodiversity and small-scale farmers. *Ecological Economics*, *140*, 136–145. <https://doi.org/10.1016/j.ecolecon.2017.04.019>
- Jones, J., Zhang, E., Tucker, D., Rietz, D., Dahlbeck, D., Gomez, M., Garcia, C., Marelli, J.-P., Livingstone, D., Schnell, R., Staskawicz, B., & Cho, M.-J. (2022). Screening of cultivars for tissue culture response and establishment of genetic transformation in a high-yielding and disease-resistant cultivar of *Theobroma cacao*. *In Vitro Cellular & Developmental Biology – Plant*, *58*(1), 133–145. <https://doi.org/10.1007/s11627-021-10205-0>
- Kadow, D., Bohlmann, J., Phillips, W., & Lieberei, R. (2013). Identification of main fine flavour components in two genotypes of the cocoa tree (*Theobroma cacao* L.). *Journal of Applied Botany and Food Quality*, *86*(1), 90–98. <https://doi.org/10.5073/JABFQ.2013.086.013>

- Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P. (2021). Global land use/land cover with Sentinel 2 and deep learning. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS* (pp. 4704–4707). IEEE. <https://doi.org/10.1109/IGARSS47720.2021.9553499>
- Khan, N., Motilal, L. A., Sukha, D. A., Bekele, F. L., Iwaro, A. D., Bidaisee, G. G., Umaharan, P., Grierson, L. H., & Zhang, D. (2008). Variability of butterfat content in cacao (*Theobroma cacao* L.): Combination and correlation with other seed-derived traits at the International Cocoa Genebank, Trinidad. *Plant Genetic Resources*, *6*(3), 175–186. <https://doi.org/10.1017/S1479262108994132>
- Kroeger, A., Bakhtary, H., Haupt, F., & Charlotte, S. (2017). *Eliminating deforestation from the cocoa supply chain*. World Bank Group. <https://openknowledge.worldbank.org/handle/10986/26549>
- Kuhn, D. N., Figueira, A., Lopes, U., Motamayor, J. C., Meerow, A. W., Cariaga, K., Freeman, B., Livingstone, D. S., & Schnell, R. J. (2010). Evaluating *Theobroma grandiflorum* for comparative genomic studies with *Theobroma cacao*. *Tree Genetics & Genomes*, *6*(5), 783–792. <https://doi.org/10.1007/s11295-010-0291-0>
- Lachenaud, P. (1997). Genetic/taxonomic structuring of the *Theobroma cacao* L. species fresh hypotheses. *INGENIC Newsletter*, *3*, 10–11. <https://www.incocoa.org/ingenic/newsletters.php>
- Lachenaud, P., Clément, D., Argout, X., Scalabrin, S., & Doaré, F. (2018). The Guiana cacao genetic group (*Theobroma cacao* L.): A new core collection in French Guiana. *Botany Letters*, *165*(2), 248–254. <https://doi.org/10.1080/23818107.2018.1465466>
- Lachenaud, P., Mooleedhar, V., & Couturier, C. (1997). Les cacaoyers spontanés de Guyane, nouvelles prospections. *Plantations, Recherche, Développement*, *4*(1), 25–32. https://agritrop.cirad.fr/388969/1/document_388969.pdf
- Lachenaud, P., & Sallée, B. (1993). Les cacaoyers spontanés de Guyane. Localisation, écologie et morphologie. *Café, Cacao, Thé*, *37*, 101–104.
- Lachenaud, P., Fouet, O., Couturier, C., & Lanaud, C. (2016). Cacao trees (*Theobroma cacao* L.) of the “Guiana” genetic group: “Tanpok 2012” survey. *Genetic Resources and Crop Evolution*, *63*(5), 781–790. <https://doi.org/10.1007/s10722-015-0280-0>
- Lachenaud, P., Paulin, D., Ducamp, M., & Thevenin, J.-M. (2007). Twenty years of agronomic evaluation of wild cocoa trees (*Theobroma cacao* L.) from French Guiana. *Scientia Horticulturae*, *113*(4), 313–321. <https://doi.org/10.1016/j.scienta.2007.05.016>
- Laliberté, B., End, M., Cryer, N., Daymond, A., Engels, J., Bernardus Esques, A., Gilmour, M., Lachenaud, P., Phillips-Mora, W., Turnbull, C., Umaharan, P., Zhang, D., & Weise, S. (2018). Conserving and exploiting cocoa genetic resources: The key challenges. In P. Umaharan (Ed.), *Burleigh Dodds series in agricultural science. Achieving sustainable cultivation of cocoa* (pp. 19–46). Burleigh Dodds Science Publishing. <https://doi.org/10.19103/AS.2017.0021.02>
- Lanaud, C. (1986). *Rapport de prospection au Venezuela, Février 1986: AGPG:IBPGR*. Food and Agriculture Organization of the United Nations. https://www.icgd.rdg.ac.uk/icgd/ref_notes.php?refcode=LAN86A
- Lanaud, C., Fouet, O., Legavre, T., Lopes, U., Sounigo, O., Eyango, M. C., Mermaz, B., da Silva, M. R., Llor Solórzano, R. G., Argout, X., Gyapay, G., Ebaiarrey, H. E., Colonges, K., Sanier, C., Rivallan, R., Mastin, G., Cryer, N., Boccara, M., Verdeil, J.-L., Clément, D. (2017).

Deciphering the *Theobroma cacao* self-incompatibility system: From genomics to diagnostic markers for self-compatibility. *Journal of Experimental Botany*, 68(17), 4775–4790. <https://doi.org/10.1093/jxb/erx293>

Lei, G. J., Fujii-Kashino, M., Wu, D. Z., Hisano, H., Saisho, D., Deng, F., Yamaji, N., Sato, K., Zhao, F.-J., & Ma, J. F. (2020). Breeding for low cadmium barley by introgression of a Sukkula-like transposable element. *Nature Food*, 1(8), 489–499. <https://doi.org/10.1038/s43016-020-0130-x>

Levis, C., Costa, F. R. C., Bongers, F., Peña-Claros, M., Clement, C. R., Junqueira, A. B., Neves, E. G., Tamanaha, E. K., Figueiredo, F. O. G., Salomão, R. P., Castilho, C. V., Magnusson, W. E., Phillips, O. L., Guevara, J. E., Sabatier, D., Molino, J.-F., López, D. C., Mendoza, A. M., . . . Pitman, N. C. A. (2017). Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science*, 355(6328), 925–931. <https://doi.org/10.1126/science.aal0157>

Lewis, C., Lennon, A. M., Eudoxie, G., & Umaharan, P. (2018). Genetic variation in bioaccumulation and partitioning of cadmium in *Theobroma cacao* L. *The Science of the Total Environment*, 640–641, 696–703. <https://doi.org/10.1016/j.scitotenv.2018.05.365>

Li, F., Wu, B., Yan, L., Hao, C., Qin, X., Lai, J., & Song, Y. (2019). Transcriptional profiling reveals differentially expressed genes involved in lipid biosynthesis during cacao seed development. *Scientific Reports*, 9(1), 17263. <https://doi.org/10.1038/s41598-019-53959-9>

Lindt & Sprüngli. (2018). Annual Report 2017. <https://www.statista.com/statistics/819288/worldwide-chocolate-consumption-by-country>

Lisboa, D. O., Evans, H. C., Araújo, J. P. M., Elias, S. G., & Barreto, R. W. (2020). *Moniliophthora perniciosa*, the mushroom causing witches' broom disease of cacao: Insights into its taxonomy, ecology and host range in Brazil. *Fungal Biology*, 124(12), 983–1003. <https://doi.org/10.1016/j.funbio.2020.09.001>

Livingstone, D., Royaert, S., Stack, C., Mockaitis, K., May, G., Farmer, A., Sasaki, C., Schnell, R., Kuhn, D., & Motamayor, J. C. (2015). Making a chocolate chip: Development and evaluation of a 6K SNP array for *Theobroma cacao*. *DNA Research*, 22(4), 279–291. <https://doi.org/10.1093/dnares/dsv009>

Livingstone, D., Stack, C., Mustiga, G. M., Rodezno, D. C., Suarez, C., Amores, F., Feltus, F. A., Mockaitis, K., Cornejo, O. E., & Motamayor, J. C. (2017). A larger chocolate chip-development of a 15K *Theobroma cacao* L. SNP Array to create high-density linkage maps. *Frontiers in Plant Science*, 8, 2008. <https://doi.org/10.3389/fpls.2017.02008>

Loor, S., Lachenaud, P., Fouet, O., Argout, X., Peña, G., Macias, J. C., Amores, F. M., Valdez, F., Hurtado, J., & Lanaud, C. (2015). Rescue of cacao genetic resources related to the Nacional variety: Surveys in the Ecuadorian Amazon (2010-2013). *ESPAMCIENCIA*, 6(E), 7–15. https://revistasespam.espam.edu.ec/index.php/Revista_ESPAMCIENCIA/article/view/101/84

Lopes, U. V., Pires, J. L., Gramacho, K. P., & Grattapaglia, D. (2022). Genome-wide SNP genotyping as a simple and practical tool to accelerate the development of inbred lines in outbred tree species: An example in cacao (*Theobroma cacao* L.). *PLoS One*, 17(10), e0270437. <https://doi.org/10.1371/journal.pone.0270437>

- López, M. E., Ramírez, O. A., Dubón, A., Ribeiro, T. H. C., Díaz, F. J., & Chalfun-Junior, A. (2021). Sexual compatibility in cacao clones drives arrangements in the field leading to high yield. *Scientia Horticulturae*, 287, 110276. <https://doi.org/10.1016/j.scienta.2021.110276>
- Mahabir, A., Motilal, L. A., Gopaulchan, D., Ramkissoon, S., Sankar, A., & Umaharan, P. (2020). Development of a core SNP panel for cacao (*Theobroma cacao* L.) identity analysis. *Genome*, 63(2), 103–114. <https://doi.org/10.1139/gen-2019-0071>
- Marelli, J.-P., Guest, D. I., Bailey, B. A., Evans, H. C., Brown, J. K., Junaid, M., Barreto, R. W., Lisboa, D. O., & Puig, A. S. (2019). Chocolate under threat from old and new cacao diseases. *Phytopathology*, 109(8), 1331–1343. <https://doi.org/10.1094/PHYTO-12-18-0477-RVW>
- Marita, J. M., Nienhuis, J., Pires, J. L., & Aitken, W. M. (2001). Analysis of genetic diversity in *Theobroma cacao* with emphasis on witches' broom disease resistance. *Crop Science*, 41(4), 1305–1316. <https://doi.org/10.2135/cropsci2001.4141305x>
- Martin, D. A., Osen, K., Grass, I., Hölscher, D., Tschardtke, T., Wurz, A., & Kreft, H. (2020). Land-use history determines ecosystem services and conservation value in tropical agroforestry. *Conservation Letters*, 13(5), e12740. <https://doi.org/10.1111/conl.12740>
- Maxted, N., Hunter, D., & Ríos, R. O. (2020). *Plant genetic conservation*. Cambridge University Press.
- McElroy, M. S., Navarro, A. J. R., Mustiga, G., Stack, C., Gezan, S., Peña, G., Sarabia, W., Saquicela, D., Sotomayor, I., Douglas, G. M., Migicovsky, Z., Amores, F., Tarqui, O., Myles, S., & Motamayor, J. C. (2018). Prediction of cacao (*Theobroma cacao*) resistance to *Moniliophthora* spp. Diseases via genome-wide association analysis and genomic selection. *Frontiers in Plant Science*, 9, 343. <https://doi.org/10.3389/fpls.2018.00343>
- Meinhardt, L. W., Rincones, J., Bailey, B. A., Aime, M. C., Griffith, G. W., Zhang, D., & Pereira, G. A. G. (2008). *Moniliophthora perniciosa*, the causal agent of witches' broom disease of cacao: What's new from this old foe? *Molecular Plant Pathology*, 9(5), 577–588. <https://doi.org/10.1111/j.1364-3703.2008.00496.x>
- Meter, A., Atkinson, R. J., & Laliberte, B. (2019). *Cadmium in cacao from Latin America and the Caribbean: A review of research and potential mitigation solutions*. CAF. <https://scioteca.caf.com/handle/123456789/1506>
- Moore, R. E. T., Ullah, I., Oliveira, V. H., Hammond, S. J., Strekopytov, S., Tibbett, M., Dunwell, J. M., & Rehkämper, M. (2020). Cadmium isotope fractionation reveals genetic variation in Cd uptake and translocation by *Theobroma cacao* and role of natural resistance-associated macrophage protein 5 and heavy metal ATPase-family transporters. *Horticulture Research*, 7(1), 71. <https://doi.org/10.1038/s41438-020-0292-6>
- Morales, D., & Rodriguez, G. (1987). *Report on the journey to explore and collect cocoa genetic material and wild tropical fruits in Bolivia*. CIRF/IBPGR. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=BOL88A
- Morales-Cruz, A., Ali, S. S., Minio, A., Figueroa-Balderas, R., García, J. F., Kasuga, T., Puig, A. S., Marelli, J.-P., Bailey, B. A., & Cantu, D. (2020). Independent whole-genome duplications define the architecture of the genomes of the devastating West African cacao black pod pathogen *Phytophthora megakarya* and its close relative *Phytophthora palmivora*. *G3 Genes Genomes Genetics*, 10(7), 2241–2255. <https://doi.org/10.1534/g3.120.401014>

- Morris, D. (1882). Cacao: How to grow and how to cure it. Government Printing Establishment. <https://books.google.de/books?id=2Z0zAQAAMAAJ&hl=de&pg=PP9#v=onepage&q&f=false>
- Morrissey, J., Stack, J. C., Valls, R., & Motamayor, J. C. (2019). Low- cost assembly of a cacao crop genome is able to resolve complex heterozygous bubbles. *Horticulture Research*, 6, 44. <https://doi.org/10.1038/s41438-019-0125-7>
- Motamayor, J. C., Lachenaud, P., da Silva E Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). *PLoS One*, 3(10), e3311. <https://doi.org/10.1371/journal.pone.0003311>
- Motamayor, J. C., Lachenaud, P., da Silva, E., Mota, J. W., Loor, R.G., Martinez, W. J., Graham, J., Kuhn, D. N., Brown, S., & Schnell, R. J. (2010). No mas forastero: A new protocol for meaningful cacao germplasm classification. In Cocoa Producers' Alliance 16th International Cocoa Research Conference, Bali, Indonesia.
- Motamayor, J. C., Mockaitis, K., Schmutz, J., Haiminen, N., Livingstone, D., Cornejo, O., Findley, S. D., Zheng, P., Utro, F., Royaert, S., Saski, C., Jenkins, J., Podicheti, R., Zhao, M., Scheffler, B. E., Stack, J. C., Feltus, F. A., Mustiga, G. M., Amores, F., Kuhn, D. N. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biology*, 14(6), r53. <https://doi.org/10.1186/gb-2013-14-6-r53>
- Motamayor, J. C., Risterucci, A. M., Lopez, P. A., Ortiz, C. F., Moreno, A., & Lanaud, C. (2002). Cacao domestication I: The origin of the cacao cultivated by the Mayas. *Heredity*, 89(5), 380–386. <https://doi.org/10.1038/sj.hdy.6800156>
- Motilal, L. A. (2018). The role of gene banks in preserving the genetic diversity of cacao. In P. Umaharan (Ed.), *Burleigh Dodds series in agricultural science. Achieving sustainable cultivation of cocoa* (pp. 47–100). Burleigh Dodds Science Publishing. <https://doi.org/10.19103/AS.2017.0021.03>
- Mournet, P., Albuquerque, P. S. B., Alves, R. M., Silva-Werneck, J. O., Rivallan, R., Marcellino, L. H., & Clément, D. (2020). A reference high-density genetic map of *Theobroma grandiflorum* (Willd. ex Spreng) and QTL detection for resistance to witches' broom disease (*Moniliophthora perniciosa*). *Tree Genetics & Genomes*, 16(6), 89. <https://doi.org/10.1007/s11295-020-01479-3>
- Mucherino Muñoz, J. J., Melo, C. A. F., Santana Silva, R. J., Luz, E. D.M. N., & Corrêa, R. X. (2021). Structural and functional genomics of the resistance of cacao to *Phytophthora palmivora*. *Pathogens*, 10(8), 961. <https://doi.org/10.3390/pathogens10080961>
- Mustiga, G. M., Morrissey, J., Stack, J. C., DuVal, A., Royaert, S., Jansen, J., Bizzotto, C., Villela-Dias, C., Mei, L., Cahoon, E. B., Seguine, E., Marelli, J. P., & Motamayor, J. C. (2019). Identification of climate and genetic factors that control fat content and fatty acid composition of *Theobroma cacao* L. beans. *Frontiers in Plant Science*, 10, 1159. <https://doi.org/10.3389/fpls.2019.01159>
- Nair, K. P. (2021). Cocoa (*Theobroma cacao* L.). In K. P. Nair (Ed.), *Tree crops* (pp. 153–213). Springer International Publishing. https://doi.org/10.1007/978-3-030-62140-7_5

- Neves, E. G., & Heckenberger, M. J. (2019). The call of the wild: Rethinking food production in ancient Amazonia. *Annual Review of Anthropology*, 48(1), 371–388. <https://doi.org/10.1146/annurevanthro-102218-011057>
- Nieves-Orduña, H. E., Müller, M., Krutovsky, K. V., & Gailing, O. (2021). Geographic patterns of genetic variation among cacao (*Theobroma cacao* L.) populations based on chloroplast markers. *Diversity*, 13(6), 249. <https://doi.org/10.3390/d13060249>
- O'Brien, C., Hiti-Bandaralage, J., Folgado, R., Hayward, A., Lahmeyer, S., Folsom, J., & Mitter, N. (2021). Cryopreservation of woody crops: The avocado case. *Plants*, 10(5), 934. <https://doi.org/10.3390/plants10050934>
- Ofori, A., Padi, F. K., & Amoako-Attah, I. (2020). Field evaluation of cacao progenies derived from Guiana clones for yield and black pod disease resistance. *Crop Science*, 60(1), 249–261. <https://doi.org/10.1002/csc2.20101>
- Olasupo, F. O., Adewale, D. B., Aikpokpodion, P. O., Muyiwa, A. A., Bhattacharjee, R., Gutierrez, O. A., Motamayor, J. C., Schnell, R. J., Ebai, S., & Zhang, D. (2018). Genetic identity and diversity of Nigerian cacao genebank collections verified by single nucleotide polymorphisms (SNPs): A guide to field genebank management and utilization. *Tree Genetics & Genomes*, 14(2), 32. <https://doi.org/10.1007/s11295-018-1244-2>
- Oliva-Cruz, M., Mori-Culqui, P. L., Caetano, A. C., Goñas, M., VilcaValqui, N. C., & Chavez, S. G. (2021). Total fat content and fatty acid profile of fine-aroma cocoa from northeastern Peru. *Frontiers in Nutrition*, 8, 677000. <https://doi.org/10.3389/fnut.2021.677000>
- Oliveira, B., Almeida, A. A., Santos, N., & Pirovani, C. P. (2022). Tolerance strategies and factors that influence the cadmium uptake by cacao tree. *Scientia Horticulturae*, 293, 110733. <https://doi.org/10.1016/j.scienta.2021.110733>
- Orozco-Aguilar, L., López-Sampson, A., Leandro-Muñoz, M. E., Robiglio, V., Reyes, M., Bordeaux, M., Sepúlveda, N., & Somarriba, E. (2021). Elucidating pathways and discourses linking cocoa cultivation to deforestation, reforestation, and tree cover change in Nicaragua and Peru. *Frontiers in Sustainable Food Systems*, 5, 635779. <https://doi.org/10.3389/fsufs.2021.635779>
- Osorio-Guarín, J. A., Berdugo-Cely, J. A., Coronado-Silva, R. A., Baez, E., Jaimes, Y., & Yockteng, R. (2020). Genome-wide association study reveals novel candidate genes associated with productivity and disease resistance to *Moniliophthora* spp. in cacao (*Theobroma cacao* L.). *G3 Genes Genomes Genetics*, 10(5), 1713–1725. <https://doi.org/10.1534/g3.120.401153>
- Pacheco, P., Mo, K., Dudley, N., Shapiro, A., Aguilar-Amuchastegui, N., Ling, P., Anderson, C., & Marx, A. (2021). *Deforestation fronts: Drivers and responses in a changing world*. WWF.
- Paim, V. R. L. D. M., Luz, E. D. M. N., Pires, J. L., Silva, S. D. V. M., Souza, J. T., Albuquerque, P. S. B., & Santos Filho, L. P. D. (2006). Sources of resistance to *Crinipellis pernicios*a in progenies of cacao accessions collected in the Brazilian Amazon. *Scientia Agricola*, 63(6), 572–578. <https://doi.org/10.1590/S0103-90162006000600011>
- Papalexandratou, Z., & Nielsen, D. S. (2016). It's gettin' hot in here: Breeding robust yeast starter cultures for cocoa fermentation. *Trends in Microbiology*, 24(3), 168–170. <https://doi.org/10.1016/j.tim.2016.01.003>

- Paulin, D., Ducamp, M., & Lachenaud, P. (2008). New sources of resistance to *Phytophthora megakarya* identified in wild cocoa tree populations of French Guiana. *Crop Protection*, 27(7), 1143–1147. <https://doi.org/10.1016/j.cropro.2008.01.004>
- Pereira, A. S., de Almeida, A.-A. F., da Silva Branco, M. C., Costa, M. G. C., & Ahnert, D. (2017). Combining ability, heritability and genotypic relations of different physiological traits in cacao hybrids. *PloS One*, 12(6), e0178790–e0178790.
- Pereira, A. S., Lopes, U. V., Pires, J. L., Barreto, M. A., Pereira dos Santos, L., Gonzaga, R. S., Da Souza Silva, L., Jesus de Andrade, F., Solís Bonilla, J. L., & Gramacho, K. P. (2021). Quantitative resistance to witches' broom disease in progenies of different sources of resistance. *Crop Protection*, 146, 105678. <https://doi.org/10.1016/j.cropro.2021.105678>
- Peters, C. (2000). PreColumbian silviculture and indigenous management of neotropical forests. In D. Lentz (Ed.), *Imperfect balance: Landscape transformations in the Precolumbian Americas* (pp. 203–224). Columbia University Press.
- Phillips-Mora, W., Castillo, J., Krauss, U., Rodriguez, E., & Wilkinson, M. J. (2005). Evaluation of cacao (*Theobroma cacao*) clones against seven Colombian isolates of *Moniliophthora roreri* from four pathogen genetic groups. *Plant Pathology*, 54(4), 483–490. <https://doi.org/10.1111/j.1365-3059.2005.01210.x>
- Phillips-Mora, W., Aime, M. C., & Wilkinson, M. J. (2007). Biodiversity and biogeography of the cacao (*Theobroma cacao*) pathogen *Moniliophthora roreri* in tropical America. *Plant Pathology*, 56(6), 911–922. <https://doi.org/10.1111/j.1365-3059.2007.01646.x>
- Phillips-Mora, W., Arciniegas-Leal, A., & Mata-Quirós, A. (Eds.) (2017). *Generation of Cacao Clones with Durable Resistant against Frosty Pod Rot (Moniliophthora Roreri (Cif. & Par.) Evans et al.)*. <https://www.cabdirect.org/cabdirect/abstract/20203127113>
- Phillips-Mora, W., Castillo, J., Arciniegas, A., Mata, A., Sanchez, A., Leandro, M., Astorga, C., Motamayor, J., Guyton, B., Seguine, E., & Schnell, R. (2009). Overcoming the main limiting factors of cacao production in central America through the use of improved clones developed at CATIE. In Cocoa Producers' Alliance. International Cocoa Research Conference, Bali, Indonesia.
- Phillips-Mora, W., & Wilkinson, M. J. (2007). Frosty pod of cacao: A disease with a limited geographic range but unlimited potential for damage. *Plant Pathology*, 56(6), 911–922. <https://doi.org/10.1094/phyto-97-12-1644>
- Phillips-Mora, W., Arciniegas-Leal, A., Mata-Quirós, A., & Motamayor-Arias, J. C. (2012). Catálogo de clones de cacao: seleccionados por el CATIE para siembras comerciales. Serie técnica. Manual técnico no. 105. CATIE (Centro Agronómico Tropical de Investigación y Enseñanza).
- Pimenta Neto, A. A., Laranjeira, D., Pires, J. L., & Luz, E. D. M. N. (2018). Selection of cocoa progenies for resistance to witches' broom. *Tropical Plant Pathology*, 43(5), 381–388. <https://doi.org/10.1007/s40858-017-0208-7>
- Pires, J. L., Cascardo, J. C. M., Lambert, S. V., & Figueira, A. (1998). Increasing cocoa butter yield through genetic improvement of *Theobroma cacao* L.: Seed fat content variability, inheritance, and association with seed yield. *Euphytica*, 103(1), 115–121. <https://doi.org/10.1023/A:1018327411530>

- Ploetz, R. (2007). Cacao diseases: Important threats to chocolate production worldwide. *Phytopathology*, 97(12), 1634–1639. <https://doi.org/10.1094/PHYTO-97-12-1634>
- Ploetz, R. (2016). The impact of diseases on cacao production: A global overview. In B. A. Bailey & L. W. Meinhardt (Eds.), *Cacao diseases* (pp. 33–59). Springer International Publishing. https://doi.org/10.1007/978-3-319-24789-2_2
- Pound, F. J. (1938). Cacao and witches' broom disease (*Marasmius perniciosus*) of South America with notes on other species of *Theobroma*. *Archives of Cocoa Research*, 1, 26–71.
- Pound, F. J. (1943). *First report on the selection of cacao trees for resistance to witches' broom disease (Marasmius perniciosus): Report on a recent visit to the Amazon territory of Peru, September 1942–February 1943*. Trinidad and Tobago. https://www.icgd.rdg.ac.uk/icgd/ref_notes.php?refcode=POU43A
- Prance, G. T. (1973). Phytogeographic support for the theory of Pleistocene forest refuges in the Amazon Basin, based on evidence from distribution patterns in Caryocaraceae, Chrysobalanaceae, Dichapetalaceae and Lecythidaceae. *Acta Amazonica*, 3(3), 5–26.
- Puig, A., Quintanilla, W., Matsumoto, T., Keith, L., Gutierrez, O. A., & Marelli, J.-P. (2021). Effect of cacao black pod rot screening method on disease reaction determination. *Agriculture*, 11(5), 396. <https://doi.org/10.3390/agriculture11050396>
- RAISG. (2020). Amazonía bajo presión. Red Amazónica de Información Socioambiental Georreferenciada. <https://www.amazoniasocioambiental.org/es/download/amazonia-bajo-presion-2020/>
- Rice, R. A., & Greenberg, R. (2000). Cacao cultivation and the conservation of biological diversity. *AMBIO: A Journal of the Human Environment*, 29(3), 167–173. <https://doi.org/10.1579/0044-7447-29.3.167>
- Rodriguez-Medina, C., Arana, A. C., Sounigo, O., Argout, X., Alvarado, G. A., & Yockteng, R. (2019). Cacao breeding in Colombia, past, present and future. *Breeding Science*, 69(3), 373–382. <https://doi.org/10.1270/jsbbs.19011>
- Rojas, E., Zutta, B. R., Velazco, Y. K., Montoya-Zumaeta, J. G., & Salvà-Catarineu, M. (2021). Deforestation risk in the Peruvian Amazon basin. *Environmental Conservation*, 48(4), 310–319. <https://doi.org/10.1017/S0376892921000291>
- Romero-Navarro, J. A., Phillips-Mora, W., Arciniegas-Leal, A., Mata Quirós, A., Haiminen, N., Mustiga, G., Livingstone, D., van Bakel, H., Kuhn, D. N., Parida, L., Kasarskis, A., & Motamayor, J. C. (2017). Application of genome wide association and genomic prediction for improvement of cacao productivity and resistance to black and frosty pod diseases. *Frontiers in Plant Science*, 8, 1905. <https://doi.org/10.3389/fpls.2017.01905>
- Royaert, S., Jansen, J., Silva, D. V., Jesus Branco, S. M., Livingstone, D. S., Mustiga, G., Marelli, J.-P., Araújo, I. S., Corrêa, R. X., & Motamayor, J. C. (2016). Identification of candidate genes involved in witches' broom disease resistance in a segregating mapping population of *Theobroma cacao* L. in Brazil. *BMC Genomics*, 17, 107. <https://doi.org/10.1186/s12864-016-2415-x>
- Royaert, S., Phillips-Mora, W., Arciniegas Leal, A. M., Cariaga, K., Brown, J. S., Kuhn, D. N., Schnell, R. J., & Motamayor, J. C. (2011). Identification of marker-trait associations for self-compatibility in a segregating mapping population of *Theobroma cacao* L. *Tree Genetics & Genomes*, 7(6), 1159–1168. <https://doi.org/10.1007/s11295-011-0403-5>

- Ruf, F., & Schroth, G. (2004). Chocolate forests and monocultures: A historical review of cocoa growing and its conflicting role in tropical deforestation and forest conservation. In G. Schroth, A.-M. N. Izac, H. L. Vasconcelos, C. Gascon, G. A. B. Da Fonseca, & C. A. Harvey (Eds.), *Agroforestry and biodiversity conservation in tropical landscapes* (pp. 107–134). Island Press.
- Sallée, B. (1987). *Compte rendu de prospection dans les populations naturelles de Theobroma cacao L. du Haut Camopi (Guyane Française) 14/4 au 12/5/87*. IRCC/CIRAD. https://www.icgd.rdg.ac.uk/icgd/ref_notes.php?refcode=SAL87A
- Samaniego, I., Espín, S., Quiroz, J., Rosales, C., Carrillo, W., Mena, P., & García-Viguera, C. (2021). Effect of the growing area on the fat content and the fatty acid composition of Ecuadorian cocoa beans. *International Journal of Food Sciences and Nutrition*, 72(7), 901–911. <https://doi.org/10.1080/09637486.2021.1884204>
- Sanchez, P., Jaffe, K., Gonzalez, R., Garcia, L., Muller, M., & Hernandez, A. (1988). *Proyecto colección de cacao silvestre en el Amazonas Venezolano: Informe Nr 3-PR 3/11, IBPGR/88/42*. Rome. Food and Agriculture Organization of the United Nations.
- Schnell, R. J., Brown, J. S., Kuhn, D. N., Cervantes-Martinez, C., Olano, C. T., & Motamayor, J. C. (2005). Why would we breed cacao in Florida? *Proceedings of the Florida State Horticultural Society*, 118, 189–191. <https://journals.flvc.org/fshs/article/download/85997/82913>
- Schroth, G., Faria, D., Araujo, M., Bede, L., van Bael, S. A., Cassano, C. R., Oliveira, L. C., & Delabie, J. H. C. (2011). Conservation in tropical landscape mosaics: The case of the cacao landscape of southern Bahia, Brazil. *Biodiversity and Conservation*, 20(8), 1635–1654. <https://doi.org/10.1007/s10531-011-0052-x>
- Schroth, G., & Harvey, C. A. (2007). Biodiversity conservation in cocoa production landscapes: An overview. *Biodiversity and Conservation*, 16(8), 2237–2244. <https://doi.org/10.1007/s10531-007-9195-1>
- Schroth, G., Läderach, P., Martinez-Valle, A. I., & Bunn, C. (2017). From site-level to regional adaptation planning for tropical commodities: Cocoa in West Africa. *Mitigation and Adaptation Strategies for Global Change*, 22(6), 903–927. <https://doi.org/10.1007/s11027-016-9707-y>
- Schroth, G., Läderach, P., Martinez-Valle, A. I., Bunn, C., & Jassogne, L. (2016). Vulnerability to climate change of cocoa in West Africa: Patterns, opportunities and limits to adaptation. *The Science of the Total Environment*, 556, 231–241. <https://doi.org/10.1016/j.scitotenv.2016.03.024>
- Schultes, R. E. (1979). The Amazonia as a source of new economic plants. *Economic Botany*, 33(3), 259–266. <https://www.jstor.org/stable/4254073>
- Seguine, E., & Meinhardt, L. (2014). Cacao flavor through genetics Anatomy of fine flavor. *The Manufacturing Confectioner*, 65, 25–30. <https://www.ars.usda.gov/research/publications/publication/?seqNo115=311426>
- Sereno, M. L., Albuquerque, P. S. B., Vencovsky, R., & Figueira, A. (2006). Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. *Conservation Genetics*, 7(1), 13–24. <https://doi.org/10.1007/s10592-005-7568-0>

- Seyum, E. G., Bille, N. H., Abteu, W. G., Munyengwa, N., Bell, J. M., & Cros, D. (2022). Genomic selection in tropical perennial crops and plantation trees: A review. *Molecular Breeding*, 42(10). <https://doi.org/10.1007/s11032-022-01326-4>
- Soares-Filho, B. S., Nepstad, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C. A., Voll, E., McDonald, A., Lefebvre, P., & Schlesinger, P. (2006). Modelling conservation in the Amazon basin. *Nature*, 440(7083), 520–523. <https://doi.org/10.1038/nature04389>
- Somarriba, E., & Lachenaud, P. (2013). Successional cocoa agro- forests of the Amazon–Orinoco–Guiana shield. *Forests, Trees and Livelihoods*, 22(1), 51–59. <https://doi.org/10.1080/14728028.2013.770316>
- Soria, J. V. (1970). *The latest cocoa expeditions to the Amazon Basin*. Centro Agronómico Tropical de Investigación y Enseñanza – CATIE. https://www.icgd.rdg.ac.uk/icgd/ref_notes.php?refcode=SOR70A
- Statista. (2022). Size of the Chocolate Confectionary Market Worldwide from 2016 to 2026. <https://www.statista.com/>
- Steege, H. T. (2021). *The Amazon Tree Diversity Network*. <https://atdn.myspecies.info/>
- Steege, H. T., Pitman, N. C. A., Killeen, T. J., Laurance, W. F., Peres, C. A., Guevara, J. E., Salomão, R. P., Castilho, C. V., Amaral, I. L., Almeida Matos, F. D., Souza Coelho, L., Magnusson, W. E., Phillips, O. L., Andrade Lima Filho, D., Jesus Veiga Carim, M., Irumé, M. V., Martins, M. P., Molino, J.-F., Sabatier, D., Gamarra, L. V. (2015). Estimating the global conservation status of more than 15,000 Amazonian tree species. *Science Advances*, 1(10), e1500936. <https://doi.org/10.1126/sciadv.1500936>
- Steege, H. T., Pitman, N. C. A., Sabatier, D., Baraloto, C., Salomão, R. P., Guevara, J. E., Phillips, O. L., Castilho, C. V., Magnusson, W. E., Molino, J.-F., Monteagudo, A., Núñez Vargas, P., Montero, J. C., Feldpausch, T. R., Coronado, E. N. H., Killeen, T. J., Mostacedo, B., Vasquez, R., Assis, R. L., Silman, M. R. (2013). Hyperdominance in the Amazonian tree flora. *Science*, 342(6156), 1243092. <https://doi.org/10.1126/science.1243092>
- Steege, H. T., Prado, P. I., Lima, R. A. F., Pos, E., Souza Coelho, L., Andrade Lima Filho, D., Salomão, R. P., Amaral, I. L., Almeida Matos, F. D., Castilho, C. V., Phillips, O. L., Guevara, J. E., Jesus Veiga Carim, M., Cárdenas López, D., Magnusson, W. E., Wittmann, F., Martins, M. P., Sabatier, D., Irumé, M. V., Pickavance, G. (2020). Biased-corrected richness estimates for the Amazonian tree flora. *Scientific Reports*, 10(1), 10130. <https://doi.org/10.1038/s41598-020-66686-3>
- Thomas, E., van Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G., & van Etten, J. (2012). Present spatial diversity patterns of *Theobroma cacao* L. in the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. *PLoS One*, 7(10), e47676. <https://doi.org/10.1371/journal.pone.0047676>
- Thormann, I., Fiorino, E., Halewood, M., & Engels, J. M. M. (2015). Plant genetic resources collections and associated information as a baseline resource for genetic diversity studies: An assessment of the IBPGR-supported collections. *Genetic Resources and Crop Evolution*, 62(8), 1279–1293. <https://doi.org/10.1007/s10722-015-0231-9>
- Torres-Moreno, M., Torrescasana, E., Salas-Salvadó, J., & Blanch, C. (2015). Nutritional composition and fatty acids profile in cocoa beans and chocolates with different geographical origin and processing conditions. *Food Chemistry*, 166, 125–132. <https://doi.org/10.1016/j.foodchem.2014.05.141>

- Tscharntke, T., Clough, Y., Bhagwat, S. A., Buchori, D., Faust, H., Hertel, D., Hölscher, D., Juhrbandt, J., Kessler, M., Perfecto, I., Scherber, C., Schroth, G., Veldkamp, E., & Wanger, T. C. (2011). Multifunctional shade-tree management in tropical agroforestry land-scapes – A review. *The Journal of Applied Ecology*, 48(3), 619–629. <https://doi.org/10.1111/j.1365-2664.2010.01939.x>
- Tscharntke, T., Milder, J. C., Schroth, G., Clough, Y., DeClerck, F., Waldron, A., Rice, R., & Ghazoul, J. (2015). Conserving biodiversity through certification of tropical agroforestry crops at local and landscape scales. *Conservation Letters*, 8(1), 14–23. <https://doi.org/10.1111/conl.12110>
- Turnbull, C., & Hadley, P. (2021). *International Cocoa Germplasm Database (ICGD) [Online Database]*. CRA Ltd./ICE Futures Europe/University of Reading. <https://www.icgd.reading.ac.uk>
- Ueno, D., Yamaji, N., Kono, I., Huang, C. F., Ando, T., Yano, M., & Ma, J. F. (2010). Gene limiting cadmium accumulation in rice. *Proceedings of the National Academy of Sciences of the United States of America*, 107(38), 16500–16505. <https://doi.org/10.1073/pnas.1005396107>
- Ullah, I., Wang, Y., Eide, D. J., & Dunwell, J. M. (2018). Evolution, and functional analysis of natural resistance-associated macrophage proteins (NRAMPs) from *Theobroma cacao* and their role in cadmium accumulation. *Scientific Reports*, 8(1), 14412. <https://doi.org/10.1038/s41598-018-32819-y>
- van der Hammen, T., & Hooghiemstra, H. (2000). Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. *Quaternary Science Reviews*, 19(8), 725–742. [https://doi.org/10.1016/S0277-3791\(99\)00024-4](https://doi.org/10.1016/S0277-3791(99)00024-4)
- Vancutsem, C., Achard, F., Pekel, J.-F., Vieilledent, G., Carboni, S., Simonetti, D., Gallego, J., Aragão, L. E. O. C., & Nasi, R. (2021). Long-term (1990-2019) monitoring of forest cover changes in the humid tropics. *Science Advances*, 7(10), eabe1603. <https://doi.org/10.1126/sciadv.abe1603>
- Vanderschueren, R., Argüello, D., Blommaert, H., Montalvo, D., Barraza, F., Maurice, L., Schreck, E., Schulin, R., Lewis, C., Vazquez, J. L., Umaharan, P., Chavez, E., Sarret, G., & Smolders, E. (2021). Mitigating the level of cadmium in cacao products: Reviewing the transfer of cadmium from soil to chocolate bar. *Science of the Total Environment*, 781, 146779. <https://doi.org/10.1016/j.scitotenv.2021.146779>
- Vásquez-García, J., Santos-Pelaez, J. C., Malqui-Ramos, R., Vigo, C. N., Alvarado, C. W., & Bobadilla, L. G. (2022). Agromorphological characterization of cacao (*Theobroma cacao* L.) accessions from the germplasm bank of the National Institute of Agrarian Innovation, Peru. *Heliyon*, 8(10), e10888. <https://doi.org/10.1016/j.heliyon.2022.e10888>
- Vello, F., & Medeiros, A. (1966). *Expedicao botanica a Amazonia Brasileira: Comunicado ao Pessoal Technico No. 20. Co-ordenacao Technica*. CEPLAC/CEPEC. https://www.icgd.rdg.ac.uk/icgd/ref_search.php?refcode=VEL66A
- Vieira, L. R., Efraim, P., van de Walle, D., Clercq, N., & Dewettinck, K. (2015). Influence of Brazilian geographic region and organic agriculture on the composition and crystallization properties of cocoa butter. *Journal of the American Oil Chemists' Society*, 92(11–12), 1579–1592. <https://doi.org/10.1007/s11746-015-2728-y>

- Wang, B., Motilal, L. A., Meinhardt, L. W., Yin, J., & Zhang, D. (2020). Molecular characterization of a cacao germplasm collection maintained in Yunnan, China using single nucleotide polymorphism(SNP) markers. *Tropical Plant Biology*, *13*, 359–370. <https://doi.org/10.1007/s12042-020-09267-y>
- Wang, M.-R., Bi, W., Shukla, M. R., Ren, L., Hamborg, Z., Blystad, D.-R., Saxena, P. K., & Wang, Q.-C. (2021). Epigenetic and genetic integrity, metabolic stability, and field performance of cryopreserved plants. *Plants*, *10*(9), 1889. <https://doi.org/10.3390/plants10091889>
- Wessel, M., & Quist-Wessel, P. F. (2015). Cocoa production in West Africa, a review and analysis of recent developments. *NJAS – Wageningen Journal of Life Sciences*, *74–75*, 1–7. <https://doi.org/10.1016/j.njas.2015.09.001>
- Wever, J., Everaert, H., Coppieters, F., Rottiers, H., Dewettinck, K., Lefever, S., & Messens, K. (2019). The development of a novel SNP genotyping assay to differentiate cacao clones. *Scientific Reports*, *9*(1), 9512. <https://doi.org/10.1038/s41598-019-45884-8>
- Wickramasuriya, A. M., & Dunwell, J. M. (2018). Cacao biotechnology: Current status and future prospects. *Plant Biotechnology Journal*, *16*(1), 4–17. <https://doi.org/10.1111/pbi.12848>
- Winters, N. P. (2022). Evolutionary and functional genetics of disease resistance in *Theobroma cacao* and its wild relatives. [PhD]. The Pennsylvania State University. <https://plantscience.psu.edu/research/labs/guiltinan/publications/thesis>
- Wittmann, F., Schongart, J., Montero, J. C., Motzer, T., Junk, W. J., Piedade, M. T. F., Queiroz, H. L., & Worbes, M. (2006). Tree species composition and diversity gradients in white-water forests across the Amazon Basin. *Journal of Biogeography*, *33*(8), 1334–1347. <https://doi.org/10.1111/j.1365-2699.2006.01495.x>
- Wood, G., & Lass, R. A. (1985). *Cocoa* (4th ed.). Blackwell Science Ltd. <https://doi.org/10.1002/9780470698983>
- Yan, J., Wang, P., Wang, P., Yang, M., Lian, X., Tang, Z., Huang, C.-F., Salt, D. E., & Zhao, F. J. (2016). A loss-of-function allele of *OsHMA3* associated with high cadmium accumulation in shoots and grain of Japonica rice cultivars. *Plant, Cell & Environment*, *39*(9), 1941–1954. <https://doi.org/10.1111/pce.12747>
- Yang, J. Y., Scascitelli, M., Motilal, L. A., Sveinsson, S., Engels, J.M. M., Kane, N. C., Dempewolf, H., Zhang, D., Maharaj, K., & Cronk, Q. C. B. (2013). Complex origin of Trinitario-type *Theobroma cacao* (Malvaceae) from Trinidad and Tobago revealed using plastid genomics. *Tree Genetics & Genomes*, *9*(3), 829–840. <https://doi.org/10.1007/s11295-013-0601-4>
- Young, A. M. (1994). *The chocolate tree: A natural history of cacao*. Smithsonian Institution Press.
- Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I., Fouet, O., Argout, X., Guichoux, E., Salin, F., Solorzano, R. L., Bouchez, O., Vignes, H., Severts, P., Hurtado, J., Yepez, A., Grivetti, L., Blake, M., & Valdez, F. (2018). The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nature Ecology & Evolution*, *2*(12), 1879–1888. <https://doi.org/10.1038/s41559-018-0697-x>
- Zhang, D., Boccara, M., Motilal, L., Mischke, S., Johnson, E. S., Butler, D. R., Bailey, B., & Meinhardt, L. (2009). Molecular characterization of an earliest cacao (*Theobroma cacao* L.)

collection from Upper Amazon using microsatellite DNA markers. *Tree Genetics & Genomes*, 5(4), 595–607. <https://doi.org/10.1007/s11295-009-0212-2>

Zhang, D., Figueira, A., Motilal, L., Lachenaud, P., & Meinhardt, L. W. (2011). *Theobroma*. In C. Kole (Ed.), *Wild crop relatives: Genomic and breeding resources* (Vol. 6, pp. 277–296). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21201-7_13

Zhang, D., Martínez, W. J., Johnson, E. S., Somarriba, E., Phillips-Mora, W., Astorga, C., Mischke, S., & Meinhardt, L. W. (2012). Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. *Genetic Resources and Crop Evolution*, 59(2), 239–252. <https://doi.org/10.1007/s10722-011-9680-y>

Zhang, D., & Motilal, L. (2016). Origin, dispersal, and current global distribution of cacao genetic diversity. In B. A. Bailey & L. W. Meinhardt (Eds.), *Cacao diseases* (pp. 3–31). Springer International Publishing. https://doi.org/10.1007/978-3-319-24789-2_1

Zhang, Y., Maximova, S. N., & Guiltinan, M. J. (2015). Characterization of a stearyl-acyl carrier protein desaturase gene family from chocolate tree, *Theobroma cacao* L. *Frontiers in Plant Science*, 6, 239. <https://doi.org/10.3389/fpls.2015.00239>

Paper 2

**Geographic patterns of genetic variation among cacao (*Theobroma cacao*
L.) populations based on chloroplast markers**

Published

Diversity

5 June 2021

<https://doi.org/10.3390/d13060249>

Geographic Patterns of Genetic Variation among Cacao (*Theobroma cacao* L.) Populations Based on Chloroplast Markers

Helmuth E. Nieves-Orduña^{1,2}, Markus Müller¹, Konstantin V. Krutovsky^{1,2,3,4}, Oliver Gailing^{1,2*}

¹Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, 37077 Göttingen, Germany

²Center for Integrated Breeding Research, Georg-August University of Göttingen, 37077 Göttingen, Germany

³ Laboratory of Forest Genomics, Genome Research and Education Center, Institute of Fundamental Biology and Biotechnology, Siberian Federal University, 660036 Krasnoyarsk, Russia

⁴ Laboratory of Population Genetics, N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 119991 Moscow, Russia

*Correspondence: ogailin@gwdg.de.

Keywords: cacao; chloroplast haplotypes; geographic origin; chocolate; crop dispersal; SSR; microsatellite markers

Abstract

The cacao tree (*Theobroma cacao* L.) is native to the Amazon basin and widely cultivated in the tropics to produce seeds, the valuable raw material for the chocolate industry. Conservation of cacao genetic resources and their availability for breeding and production programs are vital for securing cacao supply. However, relatively little is still known about the phylogeographic structure of natural cacao populations. We studied the geographic distribution of cpDNA variation in different populations representing natural cacao stands, cacao farms in Ecuador, and breeding populations. We used six earlier published cacao chloroplast microsatellite markers to genotype 233 cacao samples. In total, 23 chloroplast haplotypes were identified. The highest variation of haplotypes was observed in western Amazonia including geographically restricted haplotypes. Two observed haplotypes were widespread across the Amazon basin suggesting long distance seed dispersal from west to east in Amazonia. Most cacao genetic groups identified earlier using nuclear SSRs are associated with specific chloroplast haplotypes. A single haplotype was common in selections representing cacao plantations in west Ecuador and reference Trinitario accessions. Our results can be used to determine the chloroplast diversity of accessions and in combination with phenotypic assessments can help to select geographically distinctive varieties for cacao breeding programs.

1 INTRODUCTION

Cacao trees (*Theobroma cacao* L.) are native to the Amazon basin, and their valuable seeds are the raw material for the chocolate industry (Bartley, 2005; Motamayor et al., 2008). Cultivated by 5–6 million small-holder farmers in Latin America, Africa, and Asia, cacao crops are vitally important for local economies in these continents (Beg et al., 2017; Rice & Greenberg, 2000). As an important tropical crop, there is a constant need to develop high-yielding and disease-resistant varieties (Bennett, 2003; Goenaga et al., 2009; Phillips-Mora et al., 2013). Thus, conservation and use of cacao genetic diversity are essential not only for sustainable cultivation in producing countries, but also for diverse and growing needs of the chocolate industry and consumers (Bartley, 2005; CacaoNet, 2012; Laliberté et al., 2018; Zhang et al., 2011; Zhang & Motilal, 2016).

Western Amazonia is considered to be a putative center of origin of cacao based on high phenotypic (Bartley, 2005; Cheesman, 1944; Cuatrecasas, 1964) and genetic (Cornejo et al., 2018; Motamayor et al., 2008; Sereno et al., 2006) diversity. This part of the Amazon is also a region of high biodiversity (Bass et al., 2010; Garcia-Davila et al., 2020; Myers et al., 2000; Hans Ter Steege et al., 2003; Tuomisto et al., 1995) that in the past served as a forest refugium favoring accumulation of high tree diversity (Prance, 1982; van der Hammen & Hooghiemstra, 2000). Western Amazonia is also considered a center of crop domestication (Clement et al., 2010; Clement, 1999; Clement et al., 2015; Meyer et al., 2012). Examples of cultivated tree species, whose probable origin of domestication is western Amazonia, include fruit trees, such as the ice cream bean tree (*Inga edulis* Mart.), tree grape (*Pourouma cecropiifolia* Mart.), caimito (*Pouteria caimito* Radlk.), and the Amazon nut tree (*Bertholletia excelsa* Bonpl.) (Levis et al., 2017). At least 29 crop species and 38 utilizable palms with probable origin in western Amazonia highlight the importance of the plant genetic resources observed in this part of the Amazon (C. R. Clement, 1999; Paniagua-Zambrana et al., 2007).

Cacao populations are organized in 10 genetic groups, some of which are called varieties (Amelonado, Contamana, Criollo, Curaray, Guianna, Iquitos, Marañon, Nacional, Nanay, and Purús), spread over the native range in South America (Motamayor et al., 2008). This classification is based on the Bayesian cluster analysis of 96 nuclear SSR markers genotyped in 952 cacao accessions selected from ex situ collections (Motamayor et al., 2008).

The cultivated cacao was traditionally classified into Criollo and Forastero (an umbrella designation for the Amazonian populations) varieties and a Trinitario variety, a natural hybrid of Criollo × Forastero (Bartley, 2005; Cheesman, 1944). However, the contribution of some of the 10 cacao groups to cultivated cacao is relatively small, which is represented mainly by three varieties: Amelonado, Criollo, and Nacional (Bartley, 2005; Bennett, 2003; Cornejo et al., 2018; Zhang et al., 2011; Zhang & Motilal, 2016). However, during domestication, the Criollo variety accumulated a large proportion of high-frequency deleterious mutations that affect fitness (Cornejo et al., 2018), necessitating the use of more diverse material for cacao breeding (Bennett, 2003; Cornejo et al., 2018; Zhang et al., 2011; Zhang & Motilal, 2016).

One of the approaches to better understand genetic resources available for cacao breeding is to characterize historical patterns of seed dispersal and the origin of cultivated populations using chloroplast (cpDNA) markers, such as chloroplast microsatellites (cpSSRs) (Yang et al., 2011). They are polymorphic markers of the chloroplast genome and typically maternally inherited in angiosperms, including cacao, and propagated via seeds. Therefore, they are used preferentially for phylogeographic analysis (Lemes et al., 2010; Weising & Gardner, 1999). Nevertheless, a set of standard nuclear SSRs (Irish et al., 2014; Motilal et al., 2011; Saunders et al., 2004) and SNPs (Jemmy et al., 2014; Ji et al., 2013; Lukman et al., 2014; Padi et al., 2015) are preferred for cacao identification in germplasm collections and breeding experiments.

Specifically, chloroplast markers are useful to understand geographic variation and dispersal of tropical trees (Caron et al., 2000; Cavers et al., 2003; Dick, 2010; Hamilton, 1999; Lemes et al., 2010). For instance, these markers revealed long distance transfer of seeds for plantation establishment in *Dalbergia sissoo* Roxb. (Pandey et al., 2004) and helped to select specific haplotypes to improve site adaptation in cultivated *Pinus armandii* Franch. (Jia et al., 2020). In cacao, the historical transfer of reproductive material for cultivation has focused on genotypes exhibiting disease resistance and flavor traits (Bartley, 2005).

Nine specific chloroplast microsatellites (cpSSR) are available for cacao genetic analysis (Yang et al., 2011). In cacao farms and localities in Trinidad and Tobago, these markers allowed the identification of seven haplotypes among Trinitarian cacao populations (Yang et al., 2013). Maternal lineages (seeds) used for cultivar development in Trinidad apparently originated from cacao populations of Central America, Peru, and Venezuela (Yang et al., 2013).

Ten haplotypes were identified by sequencing the chloroplast intergenic *trnH-psbA* spacer region of cultivated cacao trees in Soconusco, southern Mexico (Gutiérrez-López et al., 2016), where cacao cultivation has been recorded since colonial times (Coe & Coe, 2013). Gutiérrez-López et al. (2016) confirmed the introduction of a few mother trees as founder material for Soconusco plantations.

Although these studies did not include a wide range of samples within the natural distribution of cacao populations in the Amazon basin, they validated the use of chloroplast markers for managing cacao genetic resources. Here, we show how chloroplast genetic diversity is geographically distributed in Amazonia, and how it is represented in common cacao genotypes by using cpSSR markers. We hypothesize a center of cpDNA diversity in western Amazonia, considering the high genetic diversity observed in western cacao populations and

the recognized origin of the species in this region. Our objectives were to: (i) study the geographic distribution of cpDNA variation in a sample of cacao trees collected in the Amazon basin, and in samples of cultivated cacao from west Ecuador, (ii) identify cpDNA variation in breeding populations, and (iii) determine the correspondence between cacao chloroplast and nuclear DNA variation.

2. MATERIALS AND METHODS

2.1. Plant Material

Fresh leaves of 233 individual cacao trees were obtained for DNA extraction from the living collection maintained by the International Cocoa Quarantine Centre, Reading University, UK. Leaf samples of 154 cacao trees were selected based on widespread geographic origin from Amazon forests in five countries (Figure 1) (31 locations and 1–28 individuals per location, Table S1). This material corresponds to seeds or budwood collected during different expeditions in 1938–1988, aiming to obtain plant material carrying disease resistance traits and to study the species' ecology and variability (Allen, 1988; Pound, 1938; Sallée, 1987; Turnbull & Hadley, 2019; Vello & Madeiros, 1965). Although these plant expeditions took place in Amazonian forests, some collection sites possibly experienced pre-Columbian silvicultural management like ancient cacao cultivation or translocation (Barlow et al., 2012; C. R. Clement et al., 2015; Peters, 2000).

Samples of 30 individuals of cultivated cacao originated from nine different cacao farms in the coastal valley of Ecuador (1–6 individuals per location, Table S2, Figure 1). This group of germplasm is known as *Refractario cacao* and includes the progeny of approximately 80 different trees selected in 1937 for their resistance to witches' broom disease (Pound, 1938, 1943; Turnbull & Hadley, 2019; Zhang et al., 2008).

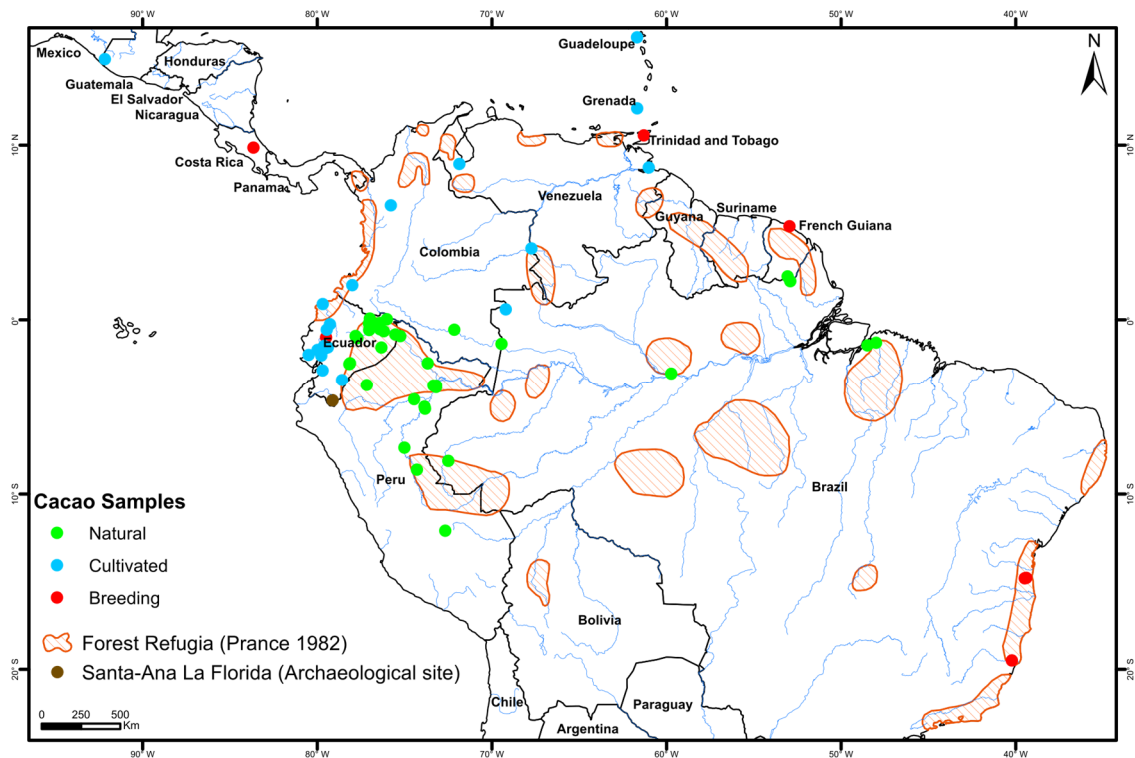


FIGURE 1. Geographic distribution of *Theobroma cacao* L. accessions in this study, forest refugial areas during the Pleistocene proposed by Prance (1982), and location of cacao use in Santa-Ana La Florida approximately 5300 years ago.

Additionally, leaf samples of 49 clones used for plantations and breeding programs were analyzed. They included the following genotypes: UF (United Fruit); ICS (Imperial Collection Selection); CRU (Cocoa Research Unit); TSH (Trinidad Selected); EET (Estación Experimental Tropical); and VB (Vassoura de Bruxa). The complete list of genotypes and their breeding origin in ten countries from Latin America and the Caribbean are presented in Supplementary Table S3.

Among the 233 samples analyzed, 83 different clones were also included in the analysis of [2] and represented different cacao genetic groups with different geographic origin: Amelonado (8), Contamana (6), Curaray (7), Guiana (2), Iquitos (23), Marañón (23), Nacional (2), and Nanay (12) (Tables S1 and S3). The International Cocoa Germplasm Database (ICGD) (Turnbull & Hadley, 2019) provides details about agronomic traits, breeding programs, plant collection expeditions and geographic origin of the 234 accessions analyzed.

2.2. DNA Extraction

To extract DNA, we used about 1 cm² of leaf tissue and the DNeasy 96 Plant Kit (Qiagen, Hilden, Germany). DNA was diluted 1:10 before PCR amplification.

2.3. Chloroplast DNA Markers

We tested the amplification and diversity of ten universal chloroplast markers (ccmps) (Weising & Gardner, 1999) in eight samples from distant regions. Additionally, we tested nine chloroplast microsatellite markers developed earlier for *T. cacao* (*CaCrSSRs*) based on the cacao chloroplast genome (Yang et al., 2011). Monomorphic amplification products were obtained with all ccmp markers, but six out of the nine cacao chloroplast markers showed clear polymorphisms, which were used then to screen the 235 samples. Four of them (*CaCr2*, *CaCr4*, *CaCr5*, and *CaCr8*) represented mononucleotide, one (*CaCr1*)—pentanucleotide, and one—octonucleotide (*CaCr9*) repeats. Although it is difficult to genotype mononucleotide repeats, it is easier to do it for chloroplast SSRs than for nuclear SSRs because they are haploid and their fragment size can be easier determined due to the lack of interference with another allele such as in diploid nuclear SSRs. Different alleles were also verified by running respective samples side-by-side during the same electrophoretic run. When the allele calling was no clear, we did repetitions to ensure a correct allele scoring.

The following six polymorphic chloroplast DNA (cpDNA) markers were amplified by PCR in two multiplexes: (1) *CaCrSSR1*, *CaCrSSR2*, and *CaCrSSR4*, and (2) *CaCrSSR5*, *CaCrSSR8*, and *CaCrSSR9*. A M13 tail (5'-CACGACGTTGTAAACGAC-3') and a PIG tail (5'-GTTTCTT-3') were attached to the 5' ends of forward and reverse primers, respectively (Kubisiak et al., 2013).

The PCR reaction mix for each primer in 14 µL volume contained: 1 µL of genomic DNA (about 0.6 ng/µL), 5.7 µL ddH₂O; 1.5 µL PCR buffer (10× Buffer B1 from Solis BioDyne,

containing Tris–HCl and (NH₄)₂SO₄), 1.5 µL MgCl₂ (25 mM), 1 µL dNTP (2.5 mM of each dNTP), 0.2 µL (5 U/µL) HOT FIREPol® Taq Polymerase from Solis BioDyne, and 0.2 µL of each forward primer, 0.5 µL of each reverse primer, and 1 µL of the M13 primer (6-FAM) for the first multiplex, but 0.3 µL and 0.75 µL of the forward and reverse primers for the CaCrSSR5 marker, respectively, 0.1 µL and 0.25 µL of the forward and reverse primers for the CaCrSSR8 marker, respectively, 0.2 µL and 0.5 µL of the forward and reverse primers for the CaCrSSR9 marker, respectively, and 1 µL M13 primer (HEX) for the second multiplex. Concentration of all primers was 5 pM/µL.

The PCR conditions for both multiplexes were 95 °C for 15 min followed by 35 cycles of 95 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min, followed by final 72 °C for 20 min, and a hold at 16 °C.

By using 1.5% agarose gel electrophoresis single bands for all PCR reactions were visualized, and the dilution ratio for the PCR product was determined. PCR products diluted at 1:10 were run on an ABI 3130xl Genetic Analyser (Applied Biosystems, Foster City, CA, USA) with GS 500 ROX used as internal size standard.

2.4. Data Analysis

Allele size analysis was performed using GeneMapper version 4.1 (Applied Biosystems, Foster City, CA, USA). Haplotypes based on all six chloroplast markers and their frequencies were determined using the Haplotype Analysis version 1.05 software (Eliades & Eliades, 2009). This software was also used to estimate total (HT) and within population (HS) haplotypic diversity, genetic differentiation (F_{ST}) between western and eastern natural populations, and between natural populations and cultivated cacao in Ecuador. The haplotype network to visualize the relationships among haplotypes was generated using the Network 5.0.1.1 software (Bandelt et al., 1999). First, a “rdf” file was created in GenAlEx 6.5 (Peakall & Smouse, 2012), and then

it was reformatted and saved in Network 5.0.1.1 as an “ych” file and used by this software as input file to generate the haplotype network using the median joining method. The created output file (“out”) was saved and used for visualization (Bandelt et al., 1999).

3. RESULTS

3.1. Identification of Haplotypes

In total, 26 alleles were identified for the six cacao chloroplast markers in 233 samples. The CaCrSSR1 marker with a pentanucleotide repeat was the most polymorphic marker with nine alleles. Six alleles were observed at CaCrSSR5, five at CaCrSSR2, three at CaCrSSR4, and two at each CaCrSSR8 and CaCrSSR9. These alleles allowed us to detect 23 haplotypes. Table 1 shows the allele compositions of these haplotypes and their frequencies. Six samples were excluded from the data analysis due to their incomplete genotyping (Tables S1–S3).

TABLE 1. *Theobroma cacao* L. chloroplast haplotypes and their frequencies observed in different locations in the Amazon, plantations, and breeding populations.

Chloroplast haplotype	CaCr1	CaCr2	CaCr4	CaCr5	CaCr8	CaCr9	Western Amazonia	Other locations*	Plantations	Breeding populations	N	Frequency, %
1	361	243	177	207	309	345	2				2	0.9
2	366	243	177	206	309	345	2	2	4	4	12	5.3
3	366	243	177	207	309	345	1				1	0.4
4	366	244	177	206	309	345			1		1	0.4
5	371	242	178	206	308	345	12			1	13	5.7
6	371	243	177	206	309	345	1	1			2	0.9
7	376	241	177	206	308	345	1				1	0.4
8	376	241	178	211	309	345			1		1	0.4
9	376	242	178	206	308	345	4		1	4	9	3.9
10	376	242	178	211	309	345	4		24	9	37	16.2
11	376	242	178	219	309	345	1				1	0.4
12	376	244	179	216	308	353	2				2	0.9
13	376	245	179	216	308	353	2				2	0.9
14	381	243	177	206	309	345	2				2	0.9
15	381	243	178	214	308	353	1				1	0.4
16	381	243	178	216	308	353	24	34	2	2	62	27.2
17	381	244	178	216	308	353	6	5	1	1	13	5.7
18	386	244	178	216	308	353	1				1	0.4
19	391	244	178	216	308	353	17		1	1	19	8.3
20	406	244	177	207	309	345	6			7	13	5.7
21	416	243	178	216	308	353				1	1	0.4
22	416	244	178	216	308	353	26	2	1	2	31	13.6
23	416	244	178	216	309	353	1				1	0.4

*Central Amazon (Manaus, Brazil) and Eastern Amazon (Belem, Brazil and Camopi, French Guyana). The allele size of each microsatellite marker (*CaCr*) is presented in base pairs. N - number of individuals.

3.2. Western Amazonia, a Center of Haplotype Diversity

We observed high cpDNA genetic diversity in western Amazonia, with 19 haplotypes detected in this region in total (Figure 2 and Figure 3). In Peruvian Amazon, 11 haplotypes were observed in three river systems, with an average of 4.6 haplotypes. The highest number of haplotypes was found along the Amazon River at Iquitos, followed by the Marañon and Nanay rivers with haplotypes H22, H16, and H19 being the most frequent, respectively. In addition, five haplotypes were observed in the south and central part of Peru along the Urubamba and Ucayali rivers (Figure 2). Nine haplotypes were observed in Ecuadorian Amazon along the Coca and Napo rivers, with haplotypes H9 and H5 being the most frequent, respectively. Finally, two haplotypes, H22 and H12, were observed in the Colombian Amazon (Figure 3). A high total haplotypic diversity ($H_T = 0.725$) and a large number of haplotypes ($N_h = 19$) were detected in western Amazonia in 28 locations (Figure 2 and Table 2). The within population haplotypic diversity was also high ($H_S = 0.676$) in this area.

TABLE 2. Chloroplast haplotype diversity within cacao populations in western and eastern Amazonia.

Amazonia	<i>N</i>	<i>N_{Loc}</i>	<i>N_h</i>	<i>H_S</i>	<i>H_T</i>	<i>F_{ST}</i>
Western	115	28	19	0.676	0.725	0.068
Eastern	32	3	2	0.037	0.087	0.572

N - sample size, *N_{Loc}* - number of locations, *N_h* - number of haplotypes, *H_S* - within population haplotypic diversity, *H_T* - total haplotypic diversity, *F_{ST}* - differentiation among populations within regions.

In contrast, only two haplotypes (H16 and H17) were found in central and eastern Amazonia (Figure 2). Both the total ($H_T = 0.087$) and the within population ($H_S = 0.037$) haplotypic diversity were low in eastern Amazonia (Table 2). When compared, eastern and western cacao populations were highly divergent ($F_{ST} = 0.500$), since they shared only two haplotypes in this study (Figure 2).

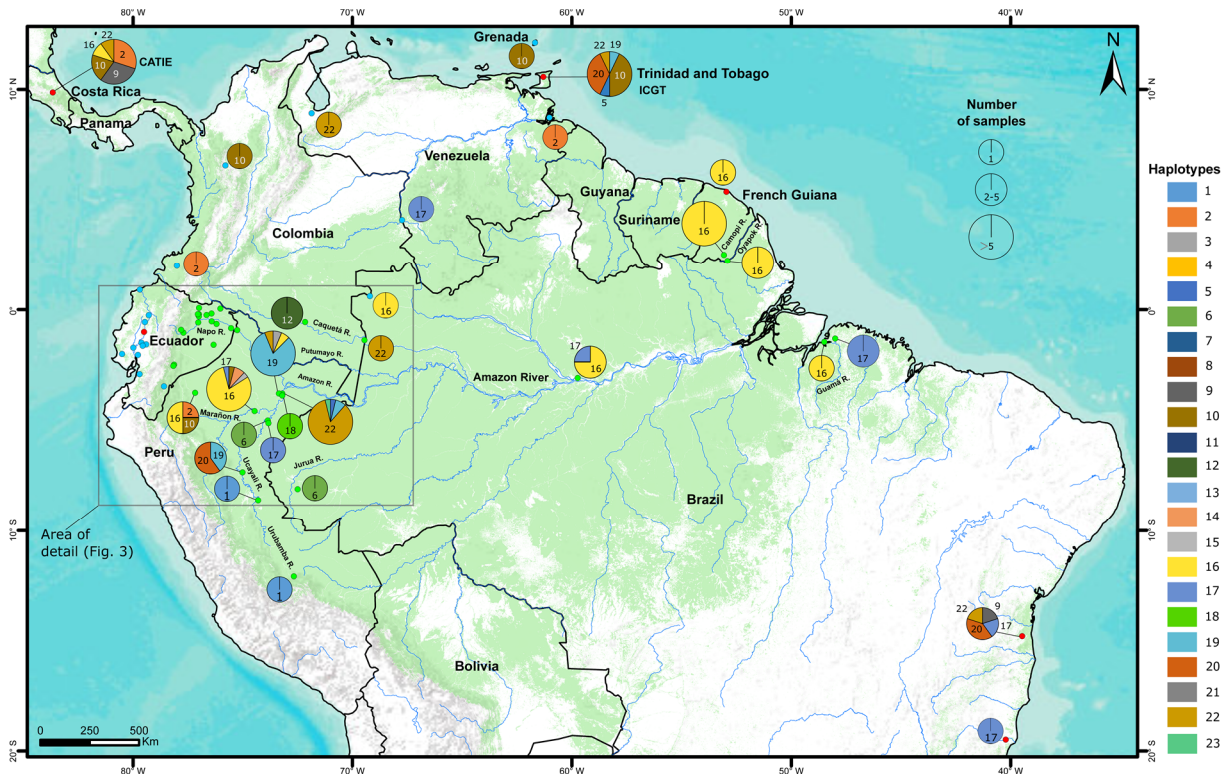


FIGURE 2. Distribution and frequencies of chloroplast haplotypes among *Theobroma cacao* L. populations in the Amazon basin. The rectangle represents the area of high haplotype diversity in northwestern Amazon (displayed in detail in Figure 3). The green dots represent samples from natural populations, the blue dots plantations, and the red dots accessions in breeding stations (CATIE-Tropical Agricultural Research and Higher Education Center, ICGT-International Cocoa Genebank Trinidad). The size of the circles is proportional to the number of samples per location.

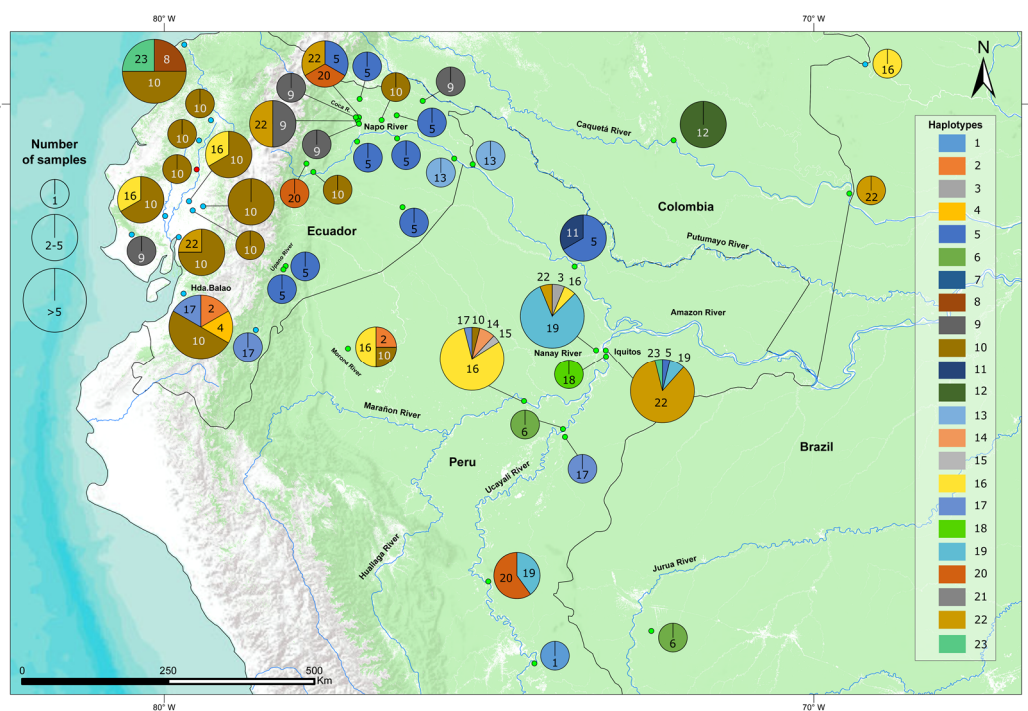


FIGURE 3. Distribution and frequencies of chloroplast haplotypes among *Theobroma cacao* L. populations in western Amazonia. The green dots represent samples from natural populations, the blue dots plantations, and the red dots accessions in breeding stations. The size of the circles is proportional to the number of samples per location.

3.3. Geographically Restricted Haplotypes in Western Amazonia

In Peru, seven geographically restricted haplotypes occurred at three locations: the confluence of the Marañon and Amazon rivers (H3, H14, H15, H18, H23), Morona River (H2) and Napo River (H11). In addition, haplotype H1 was locally common in the south of Peru along the Urubamba-Ucayali River (Figure 2).

In Ecuador, the geographically restricted haplotype H13 was observed along the Napo River. Moreover, haplotypes H5 and H9 were common in different locations along the Napo basin at the border with Colombia, and haplotype H5 was common along Upano River. Haplotype 12 was observed in the Colombian Amazon at Caquetá River (Figure 3).

3.4. Correspondence between Chloroplast and Nuclear DNA Variation

There is an agreement between the chloroplast haplotypes observed in 83 individuals (Tables S1 and S3) included in Motamayor et al. (2008) and eight cacao genetic groups. Each group generally has one dominant haplotype and other related haplotypes. Results may suggest that the cacao genetic groups proposed by Motamayor et al. (2008) are heterogeneous in terms of the chloroplast variation (Table 3). However, to study in detail patterns of correspondence between cpDNA and nuclear variation in cacao, sampling for cpDNA analysis should ideally include all individuals analyzed by Motamayor et al. (2008).

The dominant haplotypes within the genetic groups in the Peruvian Amazon were H22 in Iquitos, H19 in Nanay, H16 in Marañon, and H20 in Contamana in central Peru. In the Ecuadorian Amazon the dominant haplotype was H9 in Curaray and H10 in Nacional, and in central-eastern Amazonia H17 in Amelonado and H16 in Guiana (Table 3).

TABLE 3. Correspondence between nuclear and chloroplast DNA analyses in cacao (*Theobroma cacao* L.). Cacao genetic groups, haplotype frequency and number of samples analyzed in this study and in Motamayor et al. (2008).

Genetic group	Chloroplast haplotype (F)	N (this study)	N (Motamayor et al. 2008)
Nanay	19 (0.91)	12	121
	3 (0.08)		
Iquitos	22 (0.86)	22	75
	23 (0.04)		
	20 (0.04)		
	16 (0.04)		
Marañon	16 (0.78)	23	130
	14 (0.08)		
	17 (0.04)		
	15 (0.04)		
Guiana	14 (0.04)	2	51
	16 (1.0)		
Amelonado	17 (0.62)	8	63
	22 (0.12)		
	16 (0.12)		
	5 (0.12)		
Contamana	20 (0.51)	6	59
	1 (0.33)		
	19 (0.16)		
Curaray	9 (0.33)	6	87
	2 (0.33)		
	13 (0.16)		
	20 (0.16)		
Nacional	10 (0.50)	2	36
	6 (0.50)		

Four haplotypes were common within nuclear genetic groups: haplotype H16 was observed from west to east in the Amazon basin in Marañon (78%), Amelonado (12%), Iquitos (4%) and Guiana (100%). It is worth noticing that Marañon, Guiana, and Amelonado were also related groups based on nSSRs (Figure 2 in Motamayor et al. 2008). Haplotype H17 occurred together with the closely related H16 in Marañon demonstrating a phylogeographic pattern (Figure 4). Finally, haplotype H19 was shared between distinct clusters Nanay (91%) and Contamana (16%), and haplotype H20 was detected in the distantly related groups Contamana (50%) and Curaray (16%) (Table 3, Figure 2 in Motamayor et al. 2008).

3.5. Haplotypes in Ecuadorian Plantations

Eight haplotypes were identified in 29 samples in nine cacao plantations along the coast of Ecuador, with haplotype H10 being the most frequent one (72%) from north to southwest Ecuador (Figure 3). The geographic origin of H10 is associated with three locations in four

samples: Napo basin in Ecuador (N = 2), Morona (N = 1) and Marañon (N = 1) rivers in Peru (Figure 3).

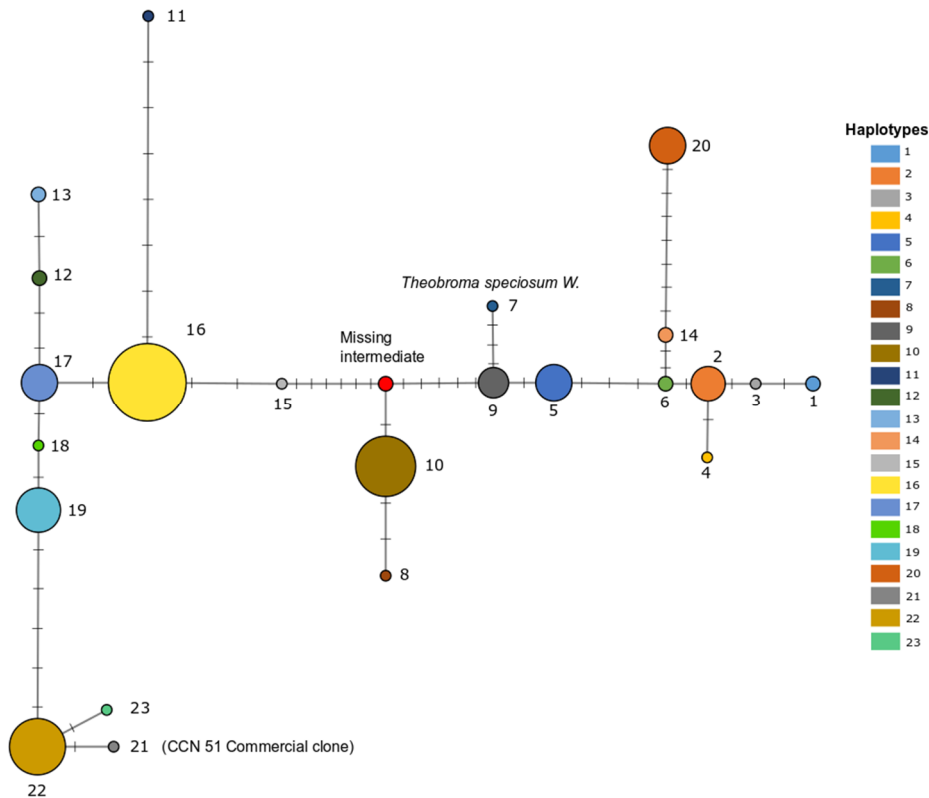


FIGURE 4. Network of chloroplast haplotypes (1–23) in *Theobroma cacao* L. populations. Each circle represents a specific haplotype, and its size reflects the respective haplotype frequency. Vertical bars in the lines connecting haplotypes indicate hypothetical mutations separating haplotypes. The distance between neighboring vertical lines is the same, but it was condensed in some lineages for better space use and visualization of the network.

Cacao farms in Ecuador also contained haplotype H2 being the second most frequent haplotype in plantations (Table 1, Figure 3). In three locations totaling three samples, H2 was observed along Napo River in Ecuador and Morona River in Peru. However, H2 and H10 may also occur in other Amazonian cacao populations that were not analyzed here.

H4 and H8 were rare haplotypes in Ecuadorian plantations observed only in Hacienda Balao and Vuelta Larga, respectively. These haplotypes were not observed in any natural population studied here (Figure 2 and Figure 3).

Low total ($H_T = 0.116$) and within population ($H_s = 0.075$) haplotypic diversity were observed in plantations of western Ecuador due to the high frequency of H10 and relatively low sample size in plantations ($N = 29$). Differentiation between cultivated and samples collected in Amazonian forest was high ($F_{ST} = 0.500$).

3.6. Haplotypes Observed in Breeding Populations and Cultivars

Cacao breeding has been mostly relying on vegetatively propagated accessions, such as Imperial College Selection (ICS), Iquitos Mixed Calabacillo (IMC), Nanay (NA), Parinari (PA), Pound, Scavina (SCA), and United Fruit (UF) (Bekele & Phillips-Mora, 2019; L.A.S. Dias, 2001; Lopes et al., 2011). Table 4 presents haplotypes detected in reference accessions and their potential geographic sources. For example, SCA 6, Pound 18, PA 139, and IMC 67 showed H5, H16, H20, and H22 haplotypes, respectively, and the geographic origin of these accessions ranged from central to northern Peru (Figure 2). In addition, cultivars, such as UF 273 (Nacional \times Amelonado), ICS, and EET (Nacional \times Unknown) revealed the common haplotype H10 observed in Ecuadorian plantations (Figure 3).

TABLE 4. Chloroplast haplotypes observed in cacao (*Theobroma cacao* L.) breeding populations, tentative geographic origin of haplotypes, source, and breeding program.

Clone name	Source	Haplo-type	Tentative geographic origin	Breeding program origin
CC 137	Cacao Center	2	Morona River, Peru	CATIE
*UF 168	United Fruit selections			UF Company, Costa Rica
NA 399	Nanay	3	Nanay River, Peru	ICG,T
ICS 68	Imperial Collection Selection	5	Western Amazonia Iquitos, Peru	ICG,T
POUND 18	Pound Collections			ICG,T
CC 252	Cacao Center	9	The northeast of Ecuador (Coca and San Miguel rivers)	CATIE
PMCT 93	Programa Mejoramiento de Cultivos Tropicales			CATIE
*UF 712	United Fruit selections			UF Company, Costa Rica
*UF 273	United Fruit selections	10	The northeast of Ecuador (Coca River) and north of Peru (Morona and Marañon rivers)	UF Company, Costa Rica
*ICS 5, 15, 42, 46, 48, 63	Imperial Collection Selection			ICG,T
EET 19, 95	Estación Experimental Tropical			EET, Pichilingue, Ecuador
PA 150, 169	Parinari	16	Marañon River, Peru	ICG,T
SCA 9,10	Scavina	19	Ucayali River, Peru	ICG,T
NA 33,	Nanay		Nanay River, Peru	ICG,T
*CRU 100	Cocoa Research Unit		Nanay River, Peru	ICG,T
*ICS 35, 41	Imperial Collection Selection	20	Western Amazonia	ICG,T
*TSA 654, 656,792	Trinidad Selected Amazons			ICG,T
SCA 11, 12, 6	Scavina			Ucayali River, Peru

Clone name	Source	Haplo-type	Tentative geographic origin	Breeding program origin
*ICS 10	Imperial Collection Selection	22	Western Amazonia	ICG,T
IMC 47, 60, 67	Iquitos Mixed Calabacillo		Iquitos, Peru	ICG,T

Accession name refers to an established accession name accepted at national and international level. * Trinitario cultivars; CATIE: Centro Agronómico Tropical de Investigación y Enseñanza, Costa Rica; ICGT: International Cocoa Genebank, Trinidad. Agronomic details (yield, disease resistance, and favorable traits) of these cacao clones are available in The International Cocoa Germplasm Database (ICGD) (Turnbull & Hadley, 2019).

The cacao cultivar CCN 51 was characterized by the unique haplotype H21. This haplotype is separated by one mutational step from H22 in the mononucleotide motif repeat of CaCrSSR2 (Figure 4). H22 was also found in the International Cacao Collection in Catie, including the Matina accession, which is also a common cultivar. Another unique and closely related haplotype is H23, which was observed along with haplotype 22 in Iquitos, Peru. H22 and H23 formed a separate cluster (lineage) in the haplotype network (Figure 4).

H10 was the most common haplotype (42%) in a set of Trinitario cultivars followed by H20 (35%) and other less frequent haplotypes H5, H19, and H22 (Table 4).

4. DISCUSSION

4.1. Distribution of Haplotype Diversity

High haplotype diversity was observed in western Amazonia (Figure 3). This region represents a hot spot of cacao diversity (Thomas et al., 2012; Zhang & Motilal, 2016), where cacao populations have a high haplotype diversity in the river systems of Marañón, Amazon (Iquitos), Nanay, and Uyacali in Peru, and along the Coca and Napo river systems in Ecuador (Figure 2 and Figure 3).

Our results support the hypothesis of decreasing cacao diversity from the western part to the eastern part of the range. The same differentiation pattern from west to east was observed in the Amazon basin for 200 cacao accessions representing ten genetic groups based on whole genome sequencing and multidimensional scaling analysis (Cornejo et al., 2018). Thus, the distribution of haplotype diversity is consistent with the suggested origin of the species in

western Amazonia (Bartley, 2005; Cheesman, 1944; Cuatrecasas, 1964; Motamayor et al., 2008).

Our results support the hypothesis of western Amazonia being considered as a forest refugial area (Haffer & Prance, 2001; Prance, 1982) and a center of crop genetic diversity (Clement, 1989; Clement et al., 2015), where cacao populations could have been restricted to temporarily isolated refugia during the Pleistocene (Figure 1). After the Pleistocene, suitable habitats during the Holocene allowed migration of cacao populations by human dispersal from western to central Amazonia along the Amazon River (Thomas et al., 2012). Cacao populations may have experienced selection in this center of crop diversity to some degree by domestication for fruit pulp and later were dispersed through the basin by pre-Columbian human expansion (Clement et al., 2010; Clement et al., 2015). Western Amazonia may also be a center of domestication which is reflected in the presence of predominant haplotypes H10, H20, and H22 both in breeding populations and plantations in western Amazonia (Figure 2 and Table 3).

Haplotypes that were observed only in western Amazonia (H12 in Colombia, H13 in Ecuador, and H11, H3, H2, H14, H15, H18, and H23 in Peru) probably are associated with cacao populations that experienced isolation in forest refugia when the Amazon forest was reduced during the Pleistocene (Figure 1) (Prance, 1982; Thomas et al., 2012; van der Hammen & Hooghiemstra, 2000). Another possible explanation assumes in situ occurrence due to edaphic adaptation to rich soils of refugial areas, which is common in tree species of Amazonia (Gentry, 1982, 1989).

Although the cacao samples studied here were distributed across the Amazon basin, they mainly represented the Upper Amazon (Figure 1). Thus, it is important to analyze additional cacao populations within the species range that could reveal new and distinctive haplotypes. For example, a recent forest survey of 1170 plots in Amazonia identified cacao as a common

tree in the basin; specifically, cacao populations were observed in southwestern and southern Amazonia, and French Guyana (Levis et al., 2017; Steege et al., 2013). Pre-Columbian human societies likely enriched these forest areas with cacao trees brought from the center of cacao diversity in western Amazonia (Levis et al., 2017).

The occurrence of only a single haplotype H16 in French Guiana could indicate human mediated dispersal of a few cacao individuals to this region. However, additional evidence to support human-mediated distribution of cacao in Amazonia is needed using more extensive sampling, specific studies such as radiocarbon and stable isotope analyses, and archaeological and ecological surveys looking at past human interactions with tropical tree species (Caetano-Andrade et al., 2020; Levis et al., 2018).

4.2. Chloroplast Haplotypes Match Genetic Groups Based on nSSRs

Haplotype diversity and distribution observed in our study partly mirrored the genetic clusters of the dendrogram based on 96 nSSRs (Figure 2 in [2]). For example, the Marañon, Guiana and Amelonado cluster is represented by related haplotypes 16 and 17 in the haplotype network (Table 4, Figure 4). Moreover, the cluster comprising Nanay and Iquitos is represented by the related haplotypes 19 and 22, and the closely related clusters Curaray and Nacional by the related haplotypes H10 and H9 in our haplotype network (Table 4, Figure 4).

The presence of specific haplotypes in different geographic regions and within different genetic groups suggest particular patterns of seed dispersal. For example, haplotype H20 was observed in Curaray (north of Ecuador), Iquitos (north of Peru), and Contamana (central Peru) groups (Figure 3, Table 4). Similarly, haplotype H19 was observed in the distantly related Contamana (central Peru) and Nanay (north of Peru) groups (Figure 3, Table 4). This may be explained by the presence of temporarily separated forest refugia in western Amazonia (Clement, 1989; Prance, 1982) where cacao populations were likely restricted during the

Pleistocene (Thomas et al., 2012). Later, during the Holocene cacao populations experienced habitat expansion when the climate became wetter and warmer, favorable for wide distribution in the Amazon basin (Thomas et al., 2012). Cacao habitat expansion aided by human mediated dispersal during the Holocene times may explain ample seed dispersal observed among cacao groups (Levis et al., 2017; Thomas et al., 2012).

Marañón, Guiana and Amelonado clustered together in the nSSR-based dendrogram (Motamayor et al., 2008); these populations share H16, and Amelonado and Marañón also have the related haplotype H17 in common (Table 4). These haplotypes reveal a wide distribution from western to eastern Amazonia (Figure 2) and likely experienced human dispersal in the basin. Indeed, current abundance of cacao populations in south-southwestern Amazon and French Guyana was explained partially by pre-Columbian human dispersal (Levis et al., 2017).

Cacao populations in French Guyana probably originated from local forest refugia during the Pleistocene–Holocene epochs (Lachenaud & Zhang, 2008). However, an alternative explanation of their origin may be human mediated seed dispersal from western to eastern Amazon (Levis et al., 2017; Sereno et al., 2006; Thomas et al., 2012). The observed wide distribution of haplotypes of H16 and H17 in the Amazon basin supports this hypothesis (Figure 2 and Figure 3). However, to clarify the origin of Amelonado and Guiana groups more extensive sampling across the species' geographic range is needed accompanied by historical records of cacao dispersal in the Amazon. However, difficulties may arise to identify the hypothetical origin of native Amazonian crops due to forest expansion during the Holocene (Clement et al., 2010).

In the case of the Nacional group, two main natural populations with H10 were observed: one in the Ecuadorian Amazon (Napo River) and another in the Peruvian Amazon (Morona River) (Figure 2). We consider that additional cpDNA analyses in Ecuadorian populations are

needed to provide more evidence about the origin of the Nacional variety. However, our results are in agreement with (Motamayor et al., 2008) who found 13 individuals from the Morona River clustered in the Nacional group. Furthermore, based on the nSSR genetic analysis of 65 wild individuals widespread in the Amazonian forest of Ecuador and eight individuals from the Morona River in Peru, (Loor Solorzano et al., 2012) suggested an ancestral origin of the Nacional group in the Southern region of Ecuadorian Amazon close to Morona River.

The distribution of H10 suggests a pattern of seed dispersal among cacao populations in western Amazonia, which supports the hypothesis of a shared origin of this haplotype from Ecuadorian and Peruvian populations and later their introduction to the coast of Ecuador, where they were established in pure plantations. Furthermore, chloroplast analysis of Chunchu, a cacao population from Peru that resembles the Nacional fine-flavor cacao (Eskes et al., 2018), would help to trace the geographic origin of fine-flavor cacao to the south of Peru as suggested by (Eskes et al., 2018).

4.3. Chloroplast Haplotypes and Domesticated Cacao

Combined analyses of nuclear and cpDNA variation can help to unravel the history of cacao domestication and to narrow down the origin of domesticated varieties. For example, the Criollo variety (not included in the present study) was domesticated from a fraction of the ancestral Curaray population about 3600 years ago (Cornejo et al., 2018). Within the Curaray population we observed haplotypes H9, H13, and H20 in northern Ecuador (Figure 3 and Table 4). CpDNA analyses of the Criollo variety could reveal its geographic origin within the distribution range of Curaray. Indeed, the geographic origin of haplotype H9 observed with high frequency in the Curaray group (Figure 4) possibly suggests that trees from the Ecuador-Colombia border served as a source of plant material during the domestication of Criollo cacao (Figure 3).

Further questions arise such as why the domestication of cacao started with the Ecuadorian Curaray group and not with other populations. It was probably because its sweet pulp was used for the elaboration of fermented beverages, which may also have favored its selection and dispersal out of the Ecuadorian Amazon (Clement et al., 2010; Loor Solorzano et al., 2012; Zarrillo et al., 2018).

4.4. Haplotype Diversity in Plantations and Breeding Populations

The high haplotype diversity observed in natural populations contrasts with the low diversity observed in cacao plantations (Table 2). We observed a high contribution of haplotype H10 in both selections made in cacao plantations in the west of Ecuador and breeding populations such as the ICS clones of Trinidad (Table 3). In addition, H10 was also observed in Colombia and Grenada, providing further evidence for introduction events of this haplotype to other countries.

Cacao cultivation in coastal Ecuador has been recorded since the 17th century, and cacao genotypes named Nacional were selected and cultivated because of their special chocolate flavor (Bartley, 2005). Currently, new sources of genetic material are required to improve flavor, increase yield and add disease resistance traits to the Nacional cultivar (Loor Solorzano et al., 2012). We suggest that the four wild provenances associated with haplotype H10 in western Amazonia of Peru and Ecuador could be screened for agronomic traits to test their use in breeding programs, potentially adding useful traits to Nacional cultivars.

The presence of six haplotypes, with haplotype H10 being the most frequent (42%), in 15 reference Trinitario accessions suggests seed introduction events occurring probably from northeast of Ecuador and north of Peru to Trinidad (Table 3). Similar to our results, five different haplotypes were observed in 21 different reference Trinitario accessions based on nine cacao cpSSRs markers (Yang et al., 2013).

The river basins associated with geographically restricted haplotypes (Figure 2 and Figure 3) could serve as source of specific adaptations and potential new traits for cacao breeding programs, but only if these samples carry favorable agronomic traits. Examples of these areas include the Napo basin of Ecuador (H5, H9, and H13), the Caquetá River (H12) in Colombia, and Ucayali River in south and central Peru (H1).

5. CONCLUSIONS

Western Amazonia contains high haplotype diversity areas in Peru and Ecuador. Populations with at least two common haplotypes may have been dispersed by humans from this center of cacao genetic diversity to new suitable habitats following a west to east route of migration in the Amazon basin. The western Amazonia has high value for cacao conservation considering its haplotype diversity and presence of geographically unique haplotypes, which were not observed in breeding populations and cultivars yet and could be evaluated for agronomic traits in support of cacao improvement.

The cacao cpDNA haplotypes observed here can be used to determine the chloroplast diversity of accessions and select distinctive haplotypes in cacao breeding programs. Additionally, the map of cpDNA haplotypes can be used to verify the geographic origin of planting material at finer geographic scale and point to potential areas for new collections. Craft chocolate makers can use this reference map to create new markets for cacao products based on geographic origin.

6. OUTLOOK

New collections in Amazonia will expand the availability of cacao genetic resources and help us better understand the patterns of geographic genetic variation and dispersal. However, rapid land use changes and gold mining in the Amazon forest are threatening the natural habitat of wild cacao populations and eroding its genetic diversity (Sierra, 2000; Steege et al., 2013;

Zhang et al., 2011). Likewise, examples of documented cacao natural populations that may reveal new haplotypes are Upper Orinoco (Patiño, 1963), Caquetá River in Colombian Amazon and Ecuadorian Amazon (Allen, 1988), Santiago and Morona rivers in northern Peru (Arevalo-Gardini et al., 2019), Chunchu in southern Peru (Eskes et al., 2018), Beni River in Bolivia (Zhang et al., 2012), the Brazilian Amazon (Almeida et al., 2015; L.A.S Dias et al., 2003; Sereno et al., 2006), southwestern and southern Amazonia (Levis et al., 2017; Wittmann et al., 2006), and Amazonian basin (Motamayor et al., 2008). Furthermore, the analysis of additional samples like those held in international (CATIE and ICGT) and national cacao collections (see Turnbull & Hadley, 2019) likely will reveal additional chloroplast diversity.

AUTHOR CONTRIBUTIONS

Conceptualization, H.E.N.-O., K.V.K. and O.G.; methodology, H.E.N.-O., K.V.K. and O.G.; validation, H.E.N.-O., M.M., K.V.K. and O.G.; formal analysis, H.E.N.-O., investigation, H.E.N.-O., M.M., K.V.K. and O.G.; resources, H.E.N.-O., K.V.K. and O.G.; data curation, H.E.N.-O., and O.G.; writing—original draft preparation, H.E.N.-O.; writing—review and editing, H.E.N.-O., M.M., K.V.K., and O.G.; visualization, H.E.N.-O.; supervision, K.V.K. and O.G.; project administration, O.G.; funding acquisition, H.E.N.-O.; K.V.K. and O.G. All authors have read and agreed to the published version of the manuscript.

FUNDING

Helmuth E. Nieves-Orduña is funded by a PhD research fellowship by the German Academic Exchange service (Deutscher Akademischer Austauschdienst, DAAD).

DATA AVAILABILITY STATEMENT

The analyzed germplasm are available on request to The International Cocoa Quarantine Centre at the University of Reading (ICQC,R) (Turnbull & Hadley, 2019).

ACKNOWLEDGMENTS

We thank the German Academic Exchange Service-DAAD for the financial support given to H. E. Nieves-Orduña. We thank Andrew Daymond and Stella Polle of the International Cocoa Quarantine Centre at the Reading University, UK for providing the plant material, Christine Radler and Alexandra Dolynska for their assistance in the laboratory, and Evert Thomas for sharing his knowledge about cacao. We acknowledge support by the Open Access Publication Funds of the Georg-August University of Göttingen.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

7 REFERENCES

- Allen, J. B. (1988). *Geographical variation and population biology in wild Theobroma cacao* [PhD]. The University of Edinburgh, Edinburgh.
- Almeida, C., Neto, F., Silva, P., Medeiros, J., & Lima, A. (2015). Genetic resources of cacao in Rondônia: historical overview, origin and inventory (in Portuguese). *Agrotrópica*, 27(2), 93–124.
- Arevalo-Gardini, E., Meinhardt, L. W., Zuñiga, L. C., Juan, A.-G., Lambert, M., & Dapeng, Z. (2019). Genetic identity and origin of “Piura Porcelana”—a fine-flavored traditional variety of cacao (*Theobroma cacao*) from the Peruvian Amazon. *Tree Genetics & Genomes*, 15(1), 11. <https://doi.org/10.1007/s11295-019-1316-y>
- Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1), 37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036>
- Barlow, J., Gardner, T. A., Lees, A. C., Parry, L., & Peres, C. A [Carlos A.] (2012). How pristine are tropical forests? An ecological perspective on the pre-Columbian human footprint in Amazonia and implications for contemporary conservation. *Biological Conservation*, 151(1), 45–49. <https://doi.org/10.1016/j.biocon.2011.10.013>
- Bartley, B. G. D. (2005). *The genetic diversity of cacao and its utilization*. CABI Pub.
- Bass, M. S., Finer, M., Jenkins, C. N., Kreft, H., Cisneros-Heredia, D. F., McCracken, S. F., Pitman, N. C. A., English, P. H., Swing, K., Villa, G., Di Fiore, A [Anthony], Voigt, C. C., & Kunz, T. H. (2010). Global conservation significance of Ecuador’s Yasuní National Park. *PloS One*, 5(1), e8767. <https://doi.org/10.1371/journal.pone.0008767>
- Beg, M. S., Ahmad, S., Jan, K., & Bashir, K. (2017). Status, supply chain and processing of cocoa - A review. *Trends in Food Science & Technology*, 66, 108–116. <https://doi.org/10.1016/j.tifs.2017.06.007>
- Bekele, F., & Phillips-Mora, W. (2019). Cacao (*Theobroma cacao* L.) breeding. In J. M. Al-Khayri, S. M. Jain, & D. V. Johnson (Eds.), *Advances in Plant Breeding Strategies: Industrial and Food Crops* (Vol. 46, pp. 409–487). Springer International Publishing. https://doi.org/10.1007/978-3-030-23265-8_12
- Bennett, A. B. (2003). Out of the Amazon: *Theobroma cacao* enters the genomic era. *Trends in Plant Science*, 8(12), 561–563. <https://doi.org/10.1016/j.tplants.2003.10.004>
- CacaoNet. (2012). *A global strategy for the conservation and use of cacao genetic resources, as the foundation for a sustainable cocoa economy*. Montpellier, France. Bioversity International.
- Caetano-Andrade, V. L., Clement, C. R., Weigel, D., Trumbore, S., Boivin, N., Schöngart, J., & Roberts, P. (2020). Tropical trees as time capsules of anthropogenic activity. *Trends in Plant Science*, 25(4), 369–380. <https://doi.org/10.1016/j.tplants.2019.12.010>

- Caron, H., Dumas, S., Marque, G., Messier, C., Bandou, E., Petit, R. J., & Kremer, A. (2000). Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. *Molecular Ecology*, 9(8), 1089–1098. <https://doi.org/10.1046/j.1365-294x.2000.00970.x>
- Cavers, S, Navarro, C, & Lowe, A. J (2003). Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, *Cedrela odorata* L., in Mesoamerica. *Molecular Ecology*, 12(6), 1451–1460. <https://doi.org/10.1046/j.1365-294X.2003.01810.x>
- Cheesman, E. E. (1944). Notes on the nomenclature, classification and possible relationships of cacao populations. *Tropical Agriculture*, 21(8), 144–159. <https://journals.sta.uwi.edu/ta/index.asp?action=viewIssue&issueId=435>
- Clement, C, Cristo-Araújo, M. de, Coppens D’Eeckenbrugge, G., Alves Pereira, A., & Picanço-Rodrigues, D. (2010). Origin and domestication of native Amazonian crops. *Diversity*, 2(1), 72–106. <https://doi.org/10.3390/d2010072>
- Clement, C. R. (1989). A center of crop genetic diversity in western Amazonia. *BioScience*, 39(9), 624–631. <https://doi.org/10.2307/1311092>
- Clement, C. R. (1999). 1492 and the loss of Amazonian crop genetic resources. ii. Crop biogeography at contact. *Economic Botany*, 53(2), 203–216. <https://doi.org/10.1007/BF02866499>
- Clement, C. R., Denevan, W. M., Heckenberger, M. J., Junqueira, A. B., Neves, E. G., Teixeira, W. G., & Woods, W. I. (2015). The domestication of Amazonia before European conquest. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 282(1812), 20150813. <https://doi.org/10.1098/rspb.2015.0813>.
- Coe, S. D., & Coe, M. D. (2013). *The true history of chocolate* (Third edition). Thames & Hudson.
- Cornejo, O. E., Yee, M.-C., Dominguez, V., Andrews, M., Sockell, A., Strandberg, E., Livingstone, D., Stack, C., Romero, A., Umaharan, P., Royaert, S., Tawari, N. R., Ng, P., Gutierrez, O., Phillips, W., Mockaitis, K., Bustamante, C. D., & Motamayor, J. C. (2018). Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process. *Communications Biology*, 1, 167. <https://doi.org/10.1038/s42003-018-0168-6>
- Cuatrecasas, J. (1964). Cacao and its allies: a taxonomic revision of the genus *Theobroma*. *Systematic Plant Studies*, 35(6), 379–614. <https://repository.si.edu/handle/10088/27110>
- Dias, L [L.A.S], Barriga, J. P., Kageyama, P. Y., & Almeida, C. M. V. C. de (2003). Variation and its distribution in wild cacao populations from the Brazilian Amazon. *Brazilian Archives of Biology and Technology*, 46(4), 507–514. <https://doi.org/10.1590/S1516-89132003000400003>
- Dias, L [L.A.S.]. (2001). The contributions of breeding. In L. Dias (Ed.), *Genetic improvement of cacao*. Food and Agriculture Organization of the United Nations. <http://ecoport.org/ep?SearchType=earticleView&earticleId=197&page=-2#section2737>
- Dick, C. W. (2010). Phylogeography and population structure of tropical trees. *Tropical Plant Biology*, 3(1), 1–3. <https://doi.org/10.1007/s12042-009-9039-0>

Eliades, N., & Eliades, D. (2009). *Haplotype Analysis* (Version 1.05) [Computer software]. Forest Genetics and Forest Tree Breeding, Georg-August University Goettingen, Germany. <http://www.uni-goettingen.de/en/134935.html>

Eskes, A., Rodriguez, C., Cruz-Condori, D., Seguíne, E., Garcia-Carrion, L., & Lachenaud, P. (2018). Large genetic diversity for fine-flavor traits unveiled in cacao (*Theobroma cacao* L.) with special attention to the native Chuncho variety in Cusco, Peru. *Agrotrópica*, 30(3), 157–174.

Garcia-Davila, C., Aldana Gomero, D., Renno, J.-F., Diaz Soria, R., Hidalgo Pizango, G., Flores Llampazo, G., Castro-Ruiz, D., Mejia de Loayza, E., Angulo Chavez, C., Mader, M., Tysklind, N., Paredes-Villanueva, K., Del Castillo Torres, D., Degen, B., & Honorio Coronado, E. N. (2020). Molecular evidence for three genetic species of *Dipteryx* in the Peruvian Amazon. *Genetica*, 148(1), 1–11. <https://doi.org/10.1007/s10709-019-00082-2>

Gentry, A. H. (1982). Neotropical floristic diversity: Phytogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny? *Annals of the Missouri Botanical Garden*, 69(3), 557. <https://doi.org/10.2307/2399084>

Gentry, A. H. (1989). Speciation in tropical forests. In L. B. Holm-Nielsen, I. C. Nielsen, & H. Balslev (Eds.), *Tropical forests: Botanical dynamics, speciation and diversity : Symposium Papers. Editors L.B. Holm-Nielsen, I.C. Nielsen and H. Balslev* (pp. 113–134). Elsevier. <https://doi.org/10.1016/B978-0-12-353550-4.50017-4>

Goenaga, R., Irizarry, H., & Irish, B. (2009). TARS series of cacao germplasm selections. *HortScience*, 44(3), 826–827. <https://doi.org/10.21273/HORTSCI.44.3.826>

Gutiérrez-López, N., Ovando-Medina, I., Salvador-Figueroa, M., Molina-Freaner, F., Avendaño-Arrazate, C. H., & Vázquez-Ovando, A. (2016). Unique haplotypes of cacao trees as revealed by *trnH-psbA* chloroplast DNA. *PeerJ*, 4, e1855. <https://doi.org/10.7717/peerj.1855>

Haffer, J., & Prance, G. (2001). Climatic forcing of evolution in Amazonia during the Cenozoic: On the refuge theory of biotic differentiation. *Amazoniana: Limnologia Et Oecologia Regionalis Systematis Fluminis Amazonas*, 16(3/4), 579–607.

Hamilton, M. B. (1999). Tropical tree gene flow and seed dispersal. *Nature*, 401(6749), 129–130. <https://doi.org/10.1038/43597>

Irish, B. M., Rios, C. J., & Goenaga, R. (2014). *Field guide to USDA-ARS TARS plant collections*. <https://doi.org/10.1515/9781400865994-010>

Jemmy, T., Ji, K., Lyndel, M., Sue, M., Stephen, Y. O., Francis, K. P., & Dapeng, Z. (2014). Verification of genetic identity of introduced cacao germplasm in Ghana using single nucleotide polymorphism (SNP) markers. *African Journal of Biotechnology*, 13(21), 2127–2136. <https://doi.org/10.5897/AJB2013.13331>

Ji, K., Zhang, D., Motilal, L. A., Boccara, M., Lachenaud, P., & Meinhardt, L. W. (2013). Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers.

Genetic Resources and Crop Evolution, 60(2), 441–453. <https://doi.org/10.1007/s10722-012-9847-1>

Jia, Y., Milne, R. I., Zhu, J., Gao, L.-M., Zhu, G.-F., Zhao, G.-F., Liu, J., & Li, Z.-H. (2020). Evolutionary legacy of a forest plantation tree species (*Pinus armandii*): Implications for widespread afforestation. *Evolutionary Applications*, 13(10), 2646–2662. <https://doi.org/10.1111/eva.13064>

Kubisiak, T. L., Nelson, C. D., Staton, M. E., Zhebentyayeva, T., Smith, C., Olukolu, B. A., Fang, G.-C., Hebard, F. V., Anagnostakis, S., Wheeler, N., Sisco, P. H., Abbott, A. G., & Sederoff, R. R. (2013). A transcriptome-based genetic map of Chinese chestnut (*Castanea mollissima*) and identification of regions of segmental homology with peach (*Prunus persica*). *Tree Genetics & Genomes*, 9(2), 557–571. <https://doi.org/10.1007/s11295-012-0579-3>

Lachenaud, P., & Zhang, D. (2008). Genetic diversity and population structure in wild stands of cacao trees (*Theobroma cacao* L.) in French Guiana. *Annals of Forest Science*, 65(3), 310. <https://doi.org/10.1051/forest:2008011>

Lablanc, B., End, M., Cryer, N., Daymond, A., Engels, J., Bernardus Eskes, A., Gilmour, M., Lachenaud, P., Phillips-Mora, W., Turnbull, C., Umaharan, P., Zhang, D., & Weise, S. (2018). Conserving and exploiting cocoa genetic resources: the key challenges. In P. Umaharan (Ed.), *Burleigh Dodds Series in Agricultural Science. Achieving sustainable cultivation of cocoa* (pp. 19–46). Burleigh Dodds Science Publishing. <https://doi.org/10.19103/AS.2017.0021.02>

Lemes, M. R., Dick, C. W., Navarro, C., Lowe, A. J., Cavers, S., & Gribel, R. (2010). Chloroplast DNA microsatellites reveal contrasting phylogeographic structure in Mahogany (*Swietenia macrophylla* King, Meliaceae) from Amazonia and Central America. *Tropical Plant Biology*, 3(1), 40–49. <https://doi.org/10.1007/s12042-010-9042-5>

Levis, C., Costa, F. R. C., Bongers, F., Peña-Claros, M., Clement, C. R., Junqueira, A. B., Neves, E. G., Tamanaha, E. K., Figueiredo, F. O. G., Salomão, R. P., Castilho, C. V., Magnusson, W. E., Phillips, O. L., Guevara, J. E., Sabatier, D., Molino, J.-F., López, D. C., Mendoza, A. M., Pitman, N. C. A., . . . Steege, H. (2017). Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science (New York, N.Y.)*, 355(6328), 925–931. <https://doi.org/10.1126/science.aal0157>

Levis, C., Flores, B. M., Moreira, P. A., Luize, B. G., Alves, R. P., Franco-Moraes, J., Lins, J., Konings, E., Peña-Claros, M., Bongers, F., Costa, F. R. C., & Clement, C. R. (2018). How people domesticated Amazonian forests. *Frontiers in Ecology and Evolution*, 5, Article 171, 15. <https://doi.org/10.3389/fevo.2017.00171>

Loor Solorzano, R. G., Fouet, O., Lemainque, A., Pavek, S., Boccara, M., Argout, X., Amores, F., Courtois, B., Risterucci, A. M., & Lanaud, C. (2012). Insight into the wild origin, migration and domestication history of the fine flavour Nacional *Theobroma cacao* L. Variety from Ecuador. *PLoS One*, 7(11), e48438. <https://doi.org/10.1371/journal.pone.0048438>

Lopes, U. V., Monteiro, W. R., Pires, J. L., Clement, D., Yamada, M. M., & Gramacho, K. P. (2011). Cacao breeding in Bahia, Brazil: strategies and results. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 11(spe), 73–81.

- Lukman, Zhang, D., Susilo, A. W., Dinarti, D., Bailey, B., Mischke, S., & Meinhardt, L. W. (2014). Genetic identity, ancestry and parentage in farmer selections of cacao from Aceh, Indonesia revealed by Single Nucleotide Polymorphism (SNP) markers. *Tropical Plant Biology*, 7(3-4), 133–143. <https://doi.org/10.1007/s12042-014-9144-6>
- Meyer, R. S., DuVal, A. E., & Jensen, H. R. (2012). Patterns and processes in crop domestication: An historical review and quantitative analysis of 203 global food crops. *The New Phytologist*, 196(1), 29–48. <https://doi.org/10.1111/j.1469-8137.2012.04253.x>
- Motamayor, J. C., Lachenaud, P., da Silva E Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PloS One*, 3(10), e3311. <https://doi.org/10.1371/journal.pone.0003311>
- Motilal, L. A., Zhang, D., Umaharan, P., Mischke, S., Pinney, S., & Meinhardt, L. W. (2011). Microsatellite fingerprinting in the International Cocoa Genebank, Trinidad: accession and plot homogeneity information for germplasm management. *Plant Genetic Resources*, 9(3), 430–438. <https://doi.org/10.1017/S147926211100058X>
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853–858. <https://doi.org/10.1038/35002501>
- Padi, F. K., Ofori, A., Takrama, J., Djan, E., Opoku, S. Y., Dadzie, A. M., Bhattacharjee, R., Motamayor, J. C., & Zhang, D. (2015). The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. *Tree Genetics & Genomes*, 11(3), 147. <https://doi.org/10.1007/s11295-015-0875-9>
- Pandey, M., Gailing, O., Leinemann, L., & Finkeldey, R. (2004). Molecular markers provide evidence for long-distance planting material transfer during plantation establishment of *Dalbergia sissoo* Roxb. in Nepal. *Annals of Forest Science*, 61(6), 603–606. <https://doi.org/10.1051/forest:2004056>
- Paniagua-Zambrana, N., Byg, A., Svenning, J.-C., Moraes, M., Grandez, C., & Balslev, H. (2007). Diversity of palm uses in the western Amazon. *Biodiversity and Conservation*, 16(10), 2771–2787. <https://doi.org/10.1007/s10531-007-9218-y>
- Patiño, V. M. (1963). *Plantas cultivadas y animales domésticos en América equinoccial. Tomo I. Frutales*. Imprenta Departamental. <http://babel.banrepcultural.org/cdm/ref/collection/p17054coll10/id/2679>
- Peakall, R., & Smouse, P. E. (2012). Genalex 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics (Oxford, England)*, 28(19), 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Peters, C. (2000). PreColumbian silviculture and indigenous management of neotropical forests. In D. Lentz (Ed.), *Imperfect Balance: Landscape Transformations in the Precolumbian Americas* (pp. 203–224). Columbia University Press.
- Phillips-Mora, W., Arciniegas-Leal, A., Mata-Quirós, A., & Motamayor, J. C. (2013). *Catalogue of cacao clones: Selected by CATIE for commercial plantings. Technical series, technical manual: Nr. 105*. CATIE.

Pound, F. J. (1938). Cacao and witches' broom disease (*Marasmius perniciosus*) of South America with notes on other species of *Theobroma*. *Archives of Cocoa Research*, 1, 26–71.

Pound, F. J. (1943). *First report on the selection of cacao trees for resistance to witches' broom disease (Marasmius perniciosus): Report on a recent visit to the Amazon territory of Peru, September 1942–February 1943*. Trinidad and Tobago. http://www.icgd.rdg.ac.uk/icgd/ref_notes.php?refcode=POU43A

Prance, G. (1982). Forest refuges: evidence from woody angiosperms. In G. Prance (Ed.), *Biological Diversification in the Tropics* (pp. 137–158). Columbia University Press.

Rice, R. A., & Greenberg, R. (2000). Cacao cultivation and the conservation of biological diversity. *AMBIO: A Journal of the Human Environment*, 29(3), 167–173. <https://doi.org/10.1579/0044-7447-29.3.167>

Sallée, B. (1987). *Compte rendu de prospection dans les populations naturelles de Theobroma cacao L. du Haut Camopi (Guyane Française) 14/4 au 12/5/87*. Kourou, Guyane. IRCC/CIRAD. http://www.icgd.rdg.ac.uk/icgd/ref_notes.php?refcode=SAL87A

Saunders, J. A., Mischke, S., Leamy, E. A., & Hemeida, A. A. (2004). Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 110(1), 41–47. <https://doi.org/10.1007/s00122-004-1762-1>

Sereno, M. L., Albuquerque, P. S. B., Vencovsky, R., & Figueira, A. (2006). Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. *Conservation Genetics*, 7(1), 13–24. <https://doi.org/10.1007/s10592-005-7568-0>

Sierra, R. (2000). Dynamics and patterns of deforestation in the western Amazon: the Napo deforestation front, 1986–1996. *Applied Geography*, 20(1), 1–16. [https://doi.org/10.1016/S0143-6228\(99\)00014-4](https://doi.org/10.1016/S0143-6228(99)00014-4)

Steege, H. T., Pitman, N. C. A., Sabatier, D., Baraloto, C., Salomão, R. P., Guevara, J. E., Phillips, O. L., Castilho, C. V., Magnusson, W. E., Molino, J.-F., Monteagudo, A., Núñez Vargas, P., Montero, J. C., Feldpausch, T. R., Coronado, E. N. H., Killeen, T. J., Mostacedo, B., Vasquez, R., Assis, R. L., . . . Silman, M. R. (2013). Hyperdominance in the Amazonian tree flora. *Science (New York, N.Y.)*, 342(6156), 1243092. <https://doi.org/10.1126/science.1243092>

Steege, H. T., Pitman, N., Sabatier, D., Castellanos, H., van der Hout, P., Daly, D. C., Silveira, M., Phillips, O., Vasquez, R., van Andel, T., Duivenvoorden, J., Oliveira, A. A., Ek, R., Lilwah, R., Thomas, R., van Essen, J., Baider, C., Maas, P., Mori, S., . . . Morawetz, W. (2003). A spatial model of tree α -diversity and tree density for the Amazon. *Biodiversity and Conservation*, 12(11), 2255–2277. <https://doi.org/10.1023/A:1024593414624>

Thomas, E., van Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G., & van Etten, J. (2012). Present spatial diversity patterns of *Theobroma cacao* L. In the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. *PLoS One*, 7(10), e47676. <https://doi.org/10.1371/journal.pone.0047676>

Tuomisto, H., Ruokolainen, K., Kalliola, R., Linna, A., Danjoy, W., & Rodriguez, Z. (1995). Dissecting amazonian biodiversity. *Science (New York, N.Y.)*, 269(5220), 63–66. <https://doi.org/10.1126/science.269.5220.63>

Turnbull, C.J., & Hadley, P. (2019). *International Cocoa Germplasm Database (ICGD) [Online Database]*. CRA Ltd./ICE Futures Europe/University of Reading, UK. <http://www.icgd.reading.ac.uk>

van der Hammen, T., & Hooghiemstra, H. (2000). Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. *Quaternary Science Reviews*, 19(8), 725–742. [https://doi.org/10.1016/S0277-3791\(99\)00024-4](https://doi.org/10.1016/S0277-3791(99)00024-4)

Vello, F., & Madeiros, A. (1965). Expedicao botanica a Amazonica brasileira. Cacao Atualidades (Brasil). *Cacao Atualidades*, 2, 47–51.

Weising, K., & Gardner, R. C. (1999). A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome*, 42(1), 9–19. <https://doi.org/10.1139/g98-104>

Wittmann, F., Schongart, J., Montero, J. C., Motzer, T., Junk, W. J., Piedade, M. T. F., Queiroz, H. L., & Worbes, M. (2006). Tree species composition and diversity gradients in white-water forests across the Amazon Basin. *Journal of Biogeography*, 33(8), 1334–1347. <https://doi.org/10.1111/j.1365-2699.2006.01495.x>

Yang, J. Y., Scascitelli, M., Motilal, L. A., Sveinsson, S., Engels, J. M. M., Kane, N. C., Dempewolf, H., Zhang, D., Maharaj, K., & Cronk, Q. C. B. (2013). Complex origin of Trinitario-type *Theobroma cacao* (Malvaceae) from Trinidad and Tobago revealed using plastid genomics. *Tree Genetics & Genomes*, 9(3), 829–840. <https://doi.org/10.1007/s11295-013-0601-4>

Yang, J.Y., Motilal, L. A., Dempewolf, H., Maharaj, K., & Cronk, Q. C. B. (2011). Chloroplast microsatellite primers for cacao (*Theobroma cacao*) and other Malvaceae. *American Journal of Botany*, 98(12), e372-e374. <https://doi.org/10.3732/ajb.1100306>

Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I., Fouet, O., Argout, X., Guichoux, E., Salin, F., Solorzano, R. L., Bouchez, O., Vignes, H., Severt, P., Hurtado, J., Yopez, A., Grivetti, L., Blake, M., & Valdez, F. (2018). The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nature Ecology & Evolution*, 2(12), 1879–1888. <https://doi.org/10.1038/s41559-018-0697-x>

Zhang, D., Boccara, M., Motilal, L., Butler, D. R., Umaharan, P., Mischke, S., & Meinhardt, L. (2008). Microsatellite variation and population structure in the “Refractario” cacao of Ecuador. *Conservation Genetics*, 9(2), 327–337. <https://doi.org/10.1007/s10592-007-9345-8>

Zhang, D., Figueira, A., Motilal, L., Lachenaud, P., & Meinhardt, L. W. (2011). *Theobroma*. In C. Kole (Ed.), *Wild Crop Relatives: Genomic and Breeding Resources* (Vol. 6, pp. 277–296). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21201-7_13

Zhang, D., Martínez, W. J., Johnson, E. S., Somarriba, E., Phillips-Mora, W., Astorga, C., Mischke, S., & Meinhardt, L. W. (2012). Genetic diversity and spatial structure in a new

distinct *Theobroma cacao* L. population in Bolivia. *Genetic Resources and Crop Evolution*, 59(2), 239–252. <https://doi.org/10.1007/s10722-011-9680-y>

Zhang, D., & Motilal, L. (2016). Origin, dispersal, and current global distribution of cacao genetic diversity. In B. A. Bailey & L. W. Meinhardt (Eds.), *Cacao Diseases* (pp. 3–31). Springer International Publishing. https://doi.org/10.1007/978-3-319-24789-2_1

8 SUPPLEMENTARY MATERIALS

The following are available online at <https://www.mdpi.com/article/10.3390/d13060249/s1>,

Table S1: Clone name, ICQC (The International Cocoa Quarantine Centre) accession number, geographic origin, samples per location, genetic group according to Motamayor et al. 2008, and chloroplasts haplotypes observed in a sample of *Theobroma cacao* L. accessions. The haplotype column shows the order of the markers *CaCr1*, *CaCr2*, *CaCr4*, *CaCr5*, *CaCr8* and *CaCr9*, and the fragment sizes are given in base pairs (bp). Details about donor collections, geographic location of collections expeditions and agronomic traits of these cacao clones are available on the ICGD (The International Cocoa Germplasm Database) website (Turnbull & Hadley, 2019). **Table S2:** Clone name, ICQC accession number, locations, and chloroplast haplotypes of *Theobroma cacao* L. in nine farms of the pacific coast of Ecuador. The haplotype column shows the order of the markers *CaCr1*, *CaCr2*, *CaCr4*, *CaCr5*, *CaCr8* and *CaCr9*, and the fragment sizes are given in base pairs (bp). Collection and agronomic details of these cacao clones are available on the ICGD website. **Table S3:** Clones names, ICQC accession number, origin of samples, genetic group (Motamayor et al. 2008), and chloroplasts haplotypes observed in a sample of *Theobroma cacao* L. genotypes used for cultivation or breeding. The haplotype column shows the order of the markers *CaCr1*, *CaCr2*, *CaCr4*, *CaCr5*, *CaCr8* and *CaCr9*, and the fragment sizes are given in base pairs (bp). Agronomic details of these cacao clones (yield, disease resistance, and favorable traits) and breeding details are available on the ICGD website.

Paper 3

Genotyping of cacao (*Theobroma cacao* L.) germplasm resources with SNP markers linked to agronomic traits reveals signs of selection

Submitted

Tree genetics & Genomes

Genotyping of cacao (*Theobroma cacao* L.) germplasm resources with SNP markers linked to agronomic traits reveals signs of selection

Helmuth E. Nieves-Orduña^{1,2*}, Markus Müller^{1,2}, Konstantin V. Krutovsky^{1,2} and Oliver Gailing^{1,2*}

¹ Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, 37077 Göttingen, Germany

² Center for Integrated Breeding Research, Georg-August University of Göttingen, 37075 Göttingen, Germany

*Corresponding authors: hnieves@gwdg.de; ogailin@gwdg.de

ORCID

Helmuth E. Nieves-Orduña <https://orcid.org/0000-0002-9588-072X>

Markus Müller <https://orcid.org/0000-0001-9990-0719>

Konstantin V. Krutovsky <https://orcid.org/0000-0002-8819-7084>

Oliver Gailing <https://orcid.org/0000-0002-4572-2408>

Abstract

The Amazonian cacao tree or the chocolate tree (*Theobroma cacao* L.) is cultivated to produce seeds, the valuable raw material for the chocolate industry. However, cacao yield is hampered by diseases and low productivity. Cacao SNP markers associated with breeding and adaptive traits provide a genetic tool for improving selection and reducing breeding cycles. Here, we genotyped 40 published SNPs associated with disease resistance and 11 SNPs with yield traits in 346 accessions using the MassARRAY® system. Four genetic clusters were identified, and two of them were observed in high proportion in managed cacao, characterized by a high proportion of admixed individuals reflecting the man-made hybrids. One cluster overrepresented in managed cacao is associated with the preference to keep superior flavor in cultivated cacao from Criollo ancestry. Introgression of wild material collected in Peru associated with disease resistance to witches' broom disease (WBD) was also observed.

Underutilized genetic resources were observed in managed cacao such as the Guiana cacao, while previous agronomic evaluation has demonstrated its good yield potential. Yield and disease resistance traits (mainly resistance to WBD) show divergence between wild and managed cacao ($F_{ST} > 0.05$) probably reflecting selection during domestication, cultivation and breeding efforts. The identified SNPs showing divergence between wild and managed cacao can be used to build breeding populations. New collections of wild cacao followed by agronomic evaluations can broaden the genetic base of cultivated cacao, especially for such traits as disease resistance to frosty pod rot (FPR).

Keywords: Amazonia, cacao, disease resistance, germplasm resources, SNPs, tropical crop breeding.

1 INTRODUCTION

The cacao (*Theobroma cacao* L.) is a valuable crop tree commonly cultivated in the tropics to produce seeds, the raw material that sustains the global chocolate industry (Coe and Coe 2013; Dand 2011). The putative origin of cacao is western Amazonia (Upper Amazon basin), and its natural range includes the wide Amazon basin, where the primary gene pool of cacao is structured in twelve genetic clusters (Gutiérrez et al. 2021a; Motamayor et al. 2008; Nieves-Orduña et al. 2021; Nieves-Orduña et al. 2023; Sereno et al. 2006; Zhang et al. 2011, 2012). Seven of these genetic clusters were identified in natural populations in western Amazonia, at the Ecuadorian-Peruvian Amazon, whereas one cluster is dominant in eastern Amazonia and French Guiana (Cornejo et al. 2018; Motamayor et al. 2008). Further population genetic analysis on areas not yet studied can identify new clusters and potential genetic resources for breeding (Nieves-Orduña et al. 2023). Cacao populations from western Amazonia were likely spread by humans to Eastern Amazonia (Levis et al. 2017). This dispersal generated a gradient of cacao genetic diversity declining from west to east Amazonia (Cornejo et al. 2018; Motamayor et al. 2008; Nieves-Orduña et al. 2021). In addition, the influences of paleoclimates and forest refugial areas also shaped the distribution of cacao in Amazonia (Lachenaud 1997; Motamayor et al. 2008; Thomas et al. 2012).

The cacao genetic resources are conserved mainly in two international and several national collections (Bekele and Phillips-Mora 2019). Continued deforestation in wild habitats threatens wild cacao diversity if proper *in situ* conservation and new collections to enrich current *ex situ* collections are not implemented (Nieves-Orduña et al. 2023). The International Cocoa Quarantine Centre at the University of Reading (ICQC, R) in the UK facilitates the global distribution of pathogen-free plant material (Daymond 2018), but cultivated cacao represents only a fraction of the species' genetic diversity (Bennett 2003; Boza et al. 2014; Zhang et al. 2011; Zhang and Motilal 2016). The global demand for

chocolate has increased cacao cultivation, but it is hampered by diseases and low yield (Gutiérrez et al. 2016; Ploetz 2016). In addition, there is evidence of deforestation linked to cacao expansion mainly in west Africa (Hoang and Kanemoto 2021; Kalischek et al. 2023) but also increased forest cover through agroforestry (Orozco-Aguilar et al. 2021). Compared to advances in the productivity of tropical crops, such as oil palm, average global cacao production has remained low since 1961 (Morrissey et al. 2019).

Advances in cacao molecular marker assisted breeding and genomic selection can improve breeding populations and help to select high yield and disease-resistant cacao genotypes (Schnell et al. 2007). Still cacao breeding depends on the extensive genetic variation existing in wild populations present in Amazonia (Motamayor et al. 2008; Nieves-Orduña et al. 2023). Genome-wide association studies (GWAS) and quantitative trait locus (QTL) mapping in cacao have identified single nucleotide polymorphism (SNP) markers linked to resistance to the black pod rot (BPR) disease caused by *Phytophthora* spp. (Gutiérrez et al. 2021b), witches' broom disease (WBD) caused by *Moniliophthora perniciosa* (Motilal et al. 2016; Royaert et al. 2016), Ceratocystis wilt (CW) caused by *Ceratocystis cacaofunesta* (Fernandes et al. 2018), and moniliasis or frosty pod rot (FPR) caused by *Moniliophthora roreri* (Gutiérrez et al. 2021b). In addition, SNPs linked to high yield traits and sexual compatibility were identified (da Silva et al. 2016; Fernandes et al. 2020). Thus, favorable identified genetic variation can be exploited for improving the selection efficiency and reducing breeding cycles.

We present here our research aimed to characterize a diverse set of 346 reference cacao accessions representing both wild and managed cacao genotyped for 51 published SNPs associated with important cacao agronomic traits, such as disease resistance, yield and sexual compatibility, using the MassARRAY® system (Agena Bioscience, Hamburg, Germany). Specifically, our aims were to 1) describe the population structure and analyze genetic clusters in wild and managed cacao, 2) identify new genetic resources for cacao breeding based on SNP

profiles, and 3) identify signatures of selection for agronomic traits that differentiate wild and managed cacao to assist breeding of superior cacao genotypes. Using published phenotypic data, we aimed to also validate disease resistance and yield associated SNP alleles and uncover valuable cacao accessions and new genetic resources to be used in breeding efforts. The obtained SNP profiles will be reported to the International Cocoa Germplasm Database (ICGD). The accessions analyzed here are in the public domain and can be accessed by any cacao-producing country, facilitating the validation of results through breeding programs and farm conditions.

2 MATERIALS AND METHODS

Germplasm and DNA Isolation

The germplasm analyzed consisted of 346 cacao accessions collected in wild habitats and from managed cacao. The managed cacao included 168 clones representing different cultivars, selections and breeding populations from several groups such as Refractario, Estación Experimental Tropical (EET), United Fruit (UF), Imperial College Selection (ICS), Trinidad Select Hybrid (TSH), SIAL (Selecao Instituto Agronomico do Leste), Selecao Instituto do Cacau (SIC) and Tropical Agriculture Research Service (TARS-Series of cacao) (Turnbull and Hadley 2023). The wild germplasm samples included 178 accessions, representing mainly northwestern Amazonia (77% of the samples), which is considered the hot spot of cacao genetic diversity (Clement et al. 2010; Cornejo et al. 2018), and French Guiana (23%). The wild cacao germplasm represented a wide geographic distribution within the cacao primary gene pool, including samples from five countries, mostly from Peru (57%). They included accessions known as the Pound Collection composed of groups coded as IMC, MO, NA, PA, POUND, and SCA. These accessions were collected in the 1930s by Frederick J. Pound in the Peruvian Amazon while searching for genotypes resistant to WBD that are now the basis for breeding WBD-resistant cacao (Bartley 2005; Zhang et al. 2011). More recent collections included cacao

accessions coded as LCT EEN collected in the Ecuadorian Amazon (Allen 1988) and GU accessions from French Guiana (Lachenaud 2015). Although collected in the wild as seeds or budwood, these groups of accessions may have experienced human intervention such as cultivation or translocation, considering the persistent effects of pre-Columbian societies in plant domestication in Amazonia (Barlow et al. 2012; Clement et al. 2015; Levis et al. 2017). Among the accessions analyzed, 137 clones represented the ten genetic groups identified by Motamayor et al. (2018): Criollo (3), Amelonado (22), Nacional which represent traditional cultivars (2), Curaray (10), Contamana (5), Guiana (15), Iquitos (30), Marañón (32), Nanay (10) and Purus (8). The complete list of samples studied is presented in Supplementary Table S1. In addition, the ICGD (<http://www.icgd.rdg.ac.uk>) provides updated agronomic details, geographic origin, and passport data of the germplasm analyzed (Turnbull and Hadley 2023).

The 346 cacao accessions analyzed are subject to international distribution under the International Treaty on Plant Genetic Resources for Food and Agriculture (CacaoNet 2012) and were obtained thanks to the ICQC at the University of Reading, UK., and the germplasm cacao collection held in the International Center for Tropical Agriculture (CATIE) in Costa Rica. We used 1 cm² of fresh leaf tissue per sample and the DNeasy 96 Plant Kit (Qiagen, Hilden, Germany) for DNA extraction.

Selection of the SNP markers

We selected 51 cacao SNPs from the published data that are supposedly associated with important agronomic traits such as disease resistance (40 SNPs) and yield (11 SNPs) and genotyped them in wild and managed cacao accessions in our study (Table 1). This set of SNPs was identified by GWAS and QTL analysis in cacao research centers at Trinidad (Motilal et al. 2016), Costa Rica (Gutiérrez et al. 2021b) and Brazil (da Silva et al. 2016; Fernandes et al. 2018; Fernandes et al. 2020; Royaert et al. 2016). The SNP panel included eight SNPs associated with disease resistance to BPR (Gutiérrez et al. 2021b), 11 SNPs linked to WBD

resistance (Motilal et al. 2016; Royaert et al. 2016), 16 SNPs related to FPR resistance (Gutiérrez et al. 2021b), and five SNPs linked to CWC resistance (Fernandes et al. 2018). In addition, the SNP set included three SNPs associated with number of seeds (Motilal et al. 2016), four SNPs related to yield components such as dry seed weight, number of pods harvested, average yield and high pod index, respectively (Fernandes et al. 2020), and four SNPs associated with flower retention (as a measure of self-compatibility) (da Silva et al. 2016).

Table 1 Fifty-one SNPs associated with agronomic traits (disease resistance and yield) in *Theobroma cacao* L. and genotyped in the present study in 346 cacao accessions

Agronomic trait (reference)	SNP ID	Identification method and study population
Witches' broom disease, WBD (Royaert et al. 2016)	Tcm003s33466269, Tcm004s00110232, Tcm006s19715703, Tcm006s25375496, Tcm007s10302466, Tcm009s08066239, Tcm009s02031341	QTL mapping: 1) F ₁ population (459 trees) at Mars Center for Cocoa science, Bahia, Brazil; 2) TSH 1188 (resistant) × CCN 51 (tolerant)
Witches' broom disease, WBD (Motilal et al. 2016)	TcSNP720, TcSNP1230, TcSNP375, TcSNP1374	Genome wide associating mapping: 483 unique accessions from the International Cocoa Genebank Trinidad (ICGT)
Ceratocystis wilt, CW (Fernandes et al. 2018)	Tcm006s13222057, Tcm006s13371871, Tcm006s13372133, Tcm004s02243097, Tcm004s02747866	QTL mapping: 1) F ₁ population (266 trees) at Mars Center for Cocoa science, Bahia, Brazil; 2) TSH 1188 (resistance) × CCN 51 (susceptible)
Frosty pod rot (FPR) or Moniliasis (Gutiérrez et al. 2021)	Tcm004s01757744, Tcm004s01737820, Tcm007s05825365, Tcm007s04093978, Tcm008s05380821, Tcm008s05909121, Tcm009s41904987, Tcm009s41915451, Tcm010s02288537, Tcm010s02455243, Tcm002s04088162, Tcm002s04156143, Tcm009s40433054, Tcm009s40465466, Tcm010s00936105, Tcm010s00913858	
Black pod rot, BPR (Gutiérrez et al. 2021)	Tcm002s08313597, Tcm002s08269115, Tcm004s28538741, Tcm004s26962336, Tcm008s04656460, Tcm008s04576575, Tcm010s22418501, Tcm010s21842721	QTL mapping: 1) F ₁ population (179 trees) at CATIE, Costa Rica; 2) POUND 7 (moderately susceptible to FPR and resistant to BPR) × UF 273 (resistant to FPR and highly susceptible to BPR)
Dry bean weight/yield, number of pods harvested (Fernandes et al. 2020)	Tcm004s00289192, Tcm004s00615809	
Average yield (Fernandes et al. 2020)	Tcm004s01127580	
Pod index (Fernandes et al. 2020)	Tcm002s23708704	

Agronomic trait (reference)	SNP ID	Identification method and study population
Number of seeds (Motilal et al. 2016)	TcSNP1370, TcSNP697, TcSNP368	Genome wide associating mapping: 483 unique accessions from the International Cocoa Gene bank Trinidad (ICGT)
Flower setting, as a measure of self-compatibility (da Silva et al. 2016)	TcSNP1866, TcSNP1867, TcSNP1869, TcSNP1870	Genome wide associating mapping: 1) 295 unique accessions from the germplasm collections at Cacao Research Center (CEPEC), Bahia, Brazil; 2) 256 randomly selected individuals from a third breeding generation.

Genotyping with MassARRAY system

The SNPs' flanking sequences were obtained from the published data (see Table 1) and were used to design two SNPs assays for the MassARRAY® system using the Assay Design Suite V2.0 (Agena Bioscience 2015). Primer adjustment, PCR amplification, SAP treatment and iPLEX reaction were done following the instructions of the manufacturer (Agena Bioscience 2019a). Allele calling was conducted using Typer Analyzer v.5.0.2137 (Agena Bioscience 2019b). SNPs with at least 75% genotyping success call rate across all samples were retained for further analysis, the same genotyping success call rate threshold was used for all DNA samples. Thus, the final data set for further analysis included 318 DNA samples and 42 SNPs (Supplementary Tables S1, S2 and S3).

Data analysis

Analysis of molecular variance (AMOVA) and F_{ST} between wild and managed cacao, F_{ST} per locus, and principal coordinate analysis (PCoA) were performed in GenAlEx 6.5 using 999 permutations (Peakall and Smouse 2012). Analysis of population structure and admixture was done using STRUCTURE 2.3.4 (Pritchard et al. 2000) with the admixture model and correlated allele frequencies of 42 SNPs by testing from $K = 1$ to $K = 10$ subpopulations with 10 repetitions for each K . The numbers of burn-ins and iterations were 10,000 and 100,000, respectively. Structure Harvester 0.6.94 (Earl and vonHoldt 2012) was used to determine the

most likely number of clusters (K) using the K method, and obtained results were visualized using CLUMPAK (Kopelman et al. 2015).

An UPGMA dendrogram based on the pairwise Nei's standard genetic distance matrix of individuals genotyped for 42 SNPs was generated using Populations 1.2.32 (Langella 2001) with 999 bootstraps on loci and visualized using Interactive Tree of Life (iTOL) 6.7.4 (Letunic and Bork 2021).

Phenotypic trait data to validate favorable SNPs associated with agronomic traits were collected from published data and the International Cocoa Germplasm Data Base (Turnbull and Hadley 2023). Based on these published data, accessions were described as either susceptible or tolerant/resistant, while no clear distinction was made between the terms tolerant and resistant. Fisher's exact test was calculated in Statistica (StatSoft Europe GmbH, Hamburg, Germany) to compare genotype distributions of wild and managed cacao, and of resistant-tolerant and susceptible germplasm to WBD and CW. Relative genotype frequencies per SNP were calculated in Excel (Microsoft Corporation, Redmond, Washington, USA). The map showing the geographic distribution of identified cacao clusters was developed using ArcGIS software (www.esri.com).

3 RESULTS

Genetic structure of accessions observed in wild and managed cacao

Although the structure analysis did not resolve the exact ten genetic clusters identified earlier in cacao populations based on microsatellite markers (Motamayor et al. 2008), it suggests four most likely clusters ($K = 4$) for the studied accessions based on 42 SNPs (Figure 1 and Supplementary Figures S1-S2).

We observed a clear pattern of differentiation of the wild germplasm when it was arranged from west to east Amazonia. Accessions collected in Ecuadorian Amazon showed a high admixture proportion of cluster two (Figure 1A). Cluster one was observed in high proportion

in the Peruvian Amazon. The IMC, NA, and PA germplasm series collected around Iquitos, Peru have similar structure and are dominated by cluster one, but the PA series also admixed much with the fourth cluster (Figure 1B). The MO series collected in Peru have a high admixture from the third cluster, which has a high representation in managed cacao. The SCA series showed a high proportion of the second cluster. The accessions collected in French Guiana can be distinguished by a high admixture from the fourth cluster (Figure 1). No private alleles were observed in the data set.

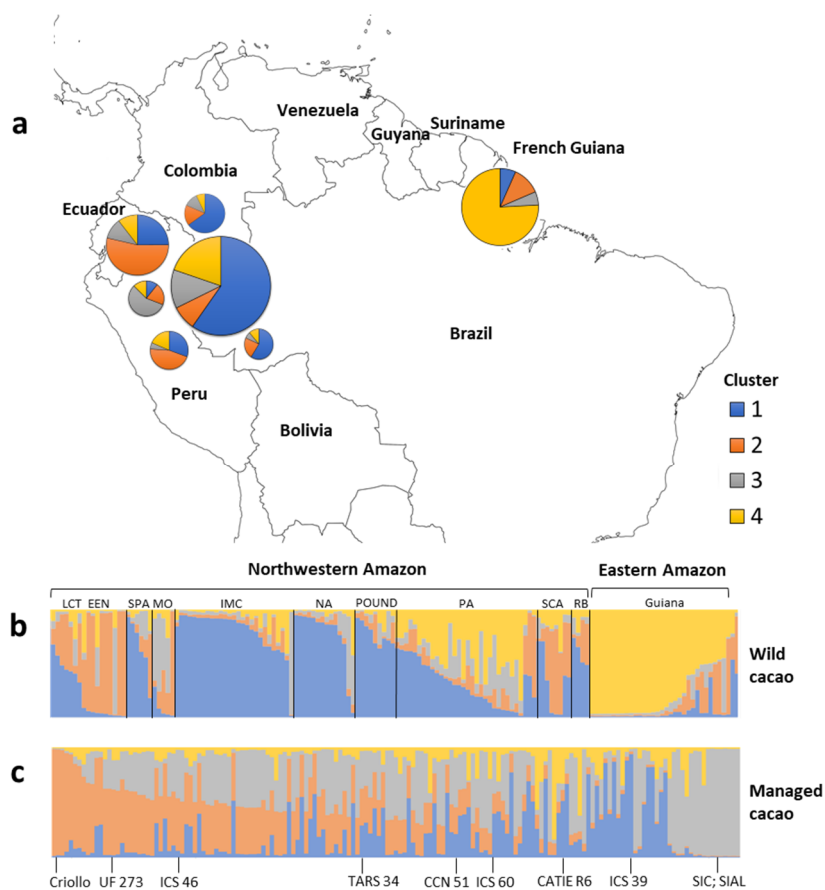


Fig. 1 Genetic admixture of four clusters identified in wild and managed cacao accessions based on 42 SNPs. **a** - geographic distribution of clusters 1- 4 in wild cacao. **b** – Q-values of individual wild cacao accessions arranged along their geographical location from west to east Amazonia. **c** - Q-values of individual managed cacao accessions arranged from high to low admixture of the second cluster. MO (Morona), IMC (Iquitos Mixed Calabacillo), NA (Nanay), POUND, PA (Parinari) and SCA (Scavina) are the names of groups of accessions from the Pound Collection

We observed that managed cacao accessions have mainly admixture from the second and third clusters, with a little admixture from the first cluster (Figure 1). A high proportion of the

second cluster was observed in the Criollo accessions (Criollo 12, 13, 65), which represent the first domesticated cacao (Cornejo et al. 2018). A higher proportion of the third cluster was observed in clones developed in Brazil, such as SIAL and SIC. In addition, we identified underrepresentation of the fourth cluster associated with Guiana accessions. Overall, managed cacao includes largely admixed accessions, reflecting hybrids in breeding populations and cultivars such as UF, ICS, TARS, CCN 51 and CATIE R6 (Figure 1 and Supplementary Figure 2B).

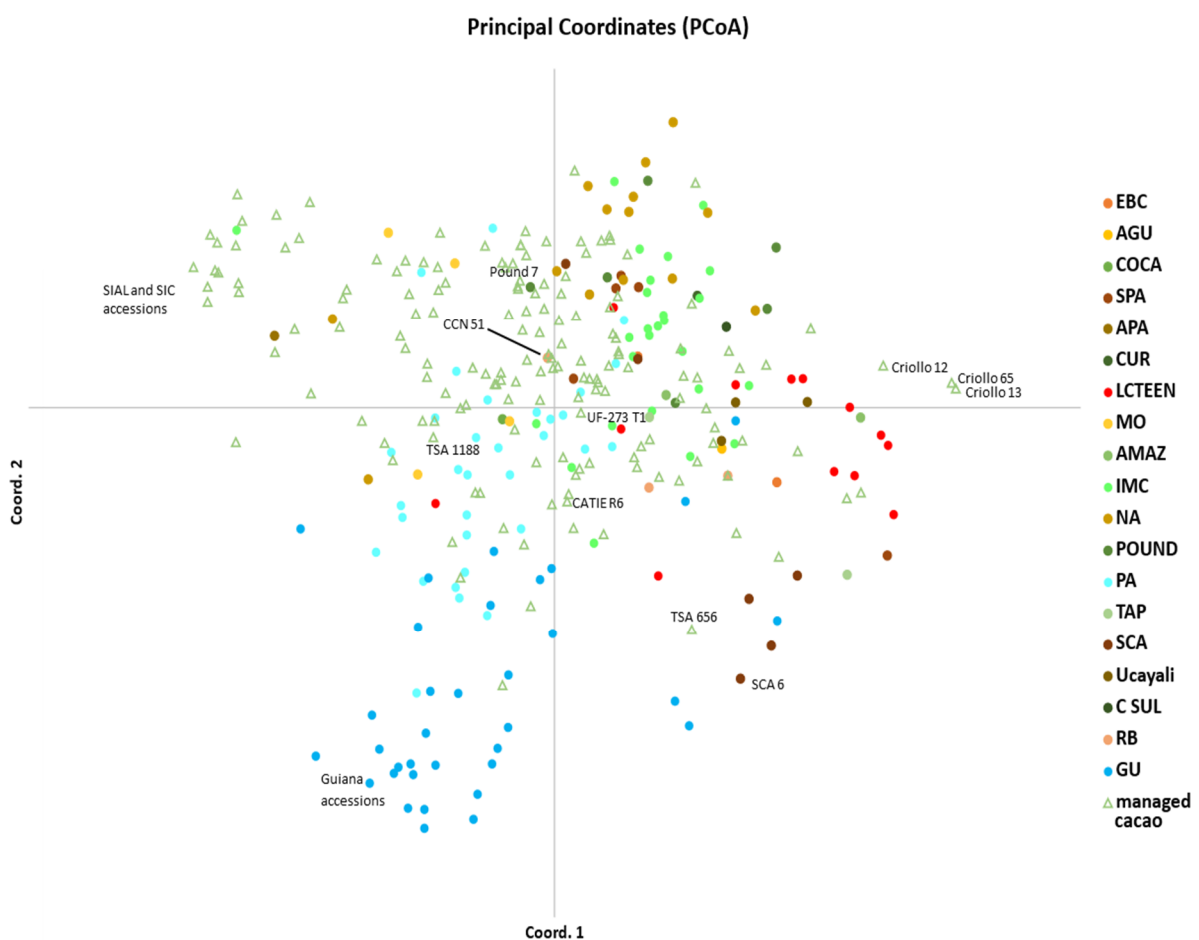


Fig. 2 Principal coordinates analysis (PCoA) of 318 cacao accessions based on genetic distance matrix estimated with 42 SNP markers linked to agronomic traits. The first component of the PCoA explains 13.09% of the total variation, and the second component explains 8.31%. Different dot colors represent different wild accessions; open green circles depict managed cacao accessions

Although there is an overlap between wild and managed cacao in the PCoA, we observed some clustering trends in wild cacao and managed cacao (Figure 2). Geographically, accessions collected in eastern Amazonia, GU-Guiana, are clustered in the lower left of the PCoA. While

wild accessions collected in western Amazonia tend to be clustered on the right side of the PCoA. Accessions such as SCA collected in central Peru are located in the lower right of the PCoA (Figure 2). The dendrogram shows a similar pattern as the PCoA, the Guiana accessions are separated as a separate group, sister to another group represented by a subset of accessions of managed cacao. The rest of the accessions of managed cacao is not well-resolved in the dendrogram (Figure 3).

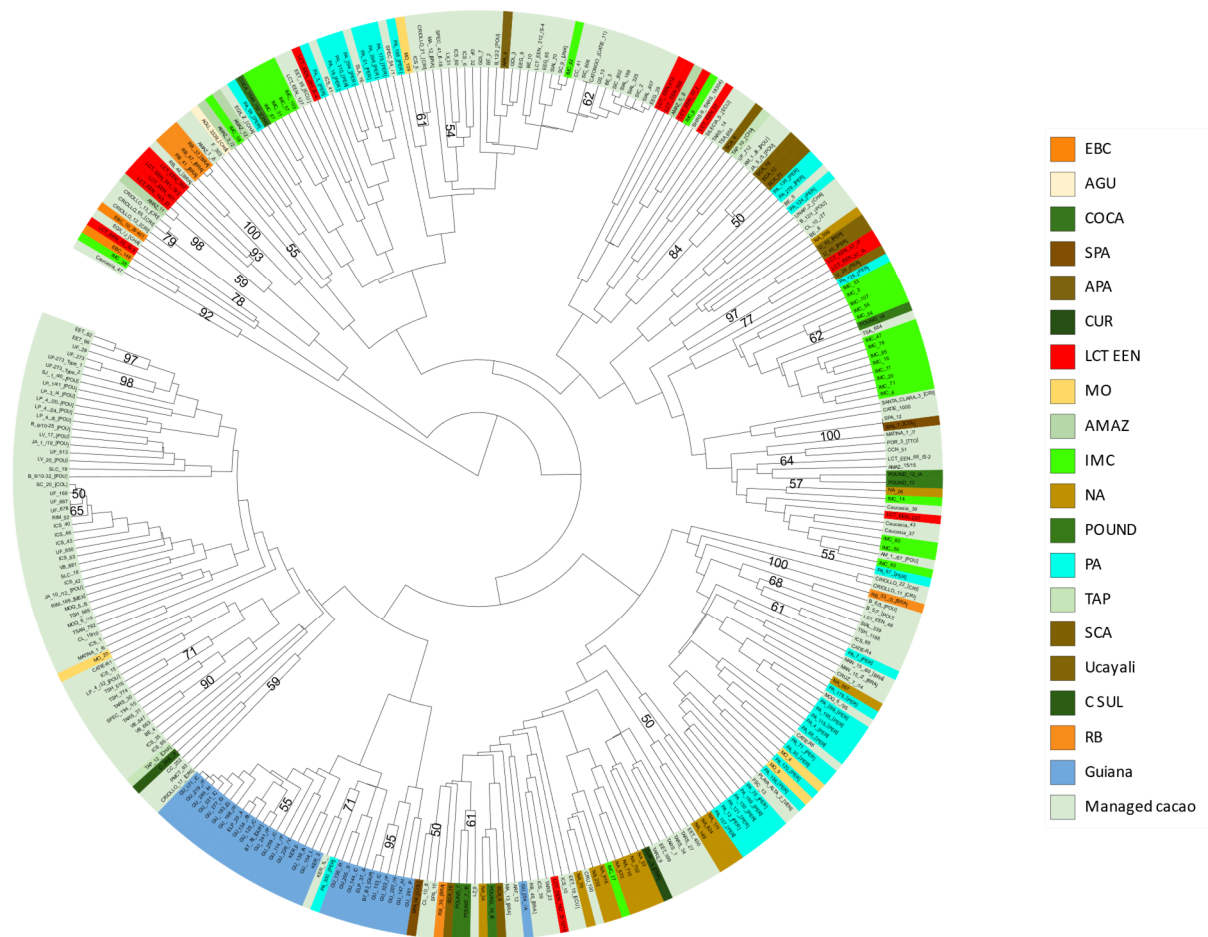


Fig. 3 UPGMA dendrogram of 318 cacao accessions based on the Nei's standard genetic distance of individuals estimated with 42 SNPs linked to agronomic traits. Clades supported by significant bootstrap values (above 50%) are shown with respective values

Managed cacao accessions tend to be clustered in the upper left of the PCoA, specially selected cacao such as SIC and SIAL developed in Bahia Brazil, but Criollo accessions (Criollo 12, 13, 65) are clustered on the right side. Within the managed cacao, cultivars and selections of high use in breeding programs such as CCN 51 and UF 273 type 1 are located in the center

of the PCoA. CCN 51 is plotted closely to RB 39, which is resistant to WBD. Wild accessions genetically close to UF 273 type 1 include PA series, such as PA 169 and PA 4. PA 4 has resistance to WBD and BPR, and PA 169 is commonly used as a source of resistance to BPR and FPR (Turnbull and Hadley 2023). The accessions IMC 6 and IMC 31 are both characterized for having a low pod index (high yield) and resistance to *Phytophthora* (Turnbull and Hadley 2023) and are closely plotted to UF 273 type 1 in the PCoA (Figure 2).

Agronomic traits showing divergence between wild and managed cacao

AMOVA analysis indicated that 5% of the total variation was caused by variation between wild and managed cacao, 37% by the variation among individuals and 58% by the variation within individuals (Table 2). In addition, 11 SNPs showed F_{ST} values above 0.05, indicating moderate genetic differentiation between managed and wild cacao. Six of these SNPs had F_{ST} values greater than 0.10, suggesting strong divergence between the two groups at these markers (Table 3).

Table 2 Analysis of molecular variance (AMOVA) in the genotyped managed and wild cacao groups

Source	d.f.	Sum of squares	Means squares	Est. var.	% of total variation	F_{ST}
Among groups	1	144.295	144.295	0.415	5	0.046*
Among individuals	316	3886.551	12.299	3.470	37	
Within individuals	318	1704.000	5.358	5.358	58	
Total	635	5734.846		9.244	100%	

* $P < 0.001$

Table 3 Eleven SNPs associated with agronomic traits showing significant divergence between managed and wild cacao ($F_{ST} > 0.05$)

Agronomic trait	SNP ID	F_{ST}	Chr.	Gene ²	Description ²
Flower setting (as a measure of self-compatibility)	TcSNP1866	0.260	4	n/a	n/a
Dry bean weight / high yield	Tcm004s00289192	0.160	4	Thecc.04G002600	Family of uncharacterized protein function
	Tcm003s33466269	0.064	3	Thecc.03G307700	RNA-binding family protein
	Tcm004s00110232	0.101	4	Thecc.04G000500	UDP-glucose pyrophosphorylase 2
Witches' broom disease (WBD)	Tcm006s19715703	0.147	6	Thecc.06G104900	Hydroxyproline-rich glycoprotein family protein
	TcSNP1230	0.082	8	Thecc.08G127800	PDI- Protein disulfide isomerase 1-1
	Tcm009s08066239	0.065	9	Thecc.09G141300	RING/U-box superfamily protein

Agro­nomic trait	SNP ID	F_{ST}	Chr.	Gene²	Description²
Ceratocystis wilt (CW)	Tcm006s13222057	0.123	6	Thecc.06G070400	Peptide deformylase 1A
Pod index ¹	Tcm002s23708704	0.107	2	Thecc.02G221200	Elongator protein 2
Moniliasis (FPR)	Tcm009s40465466	0.061	9	Thecc.09G335000	Pseudo-response regulator 7
Number of seeds	TcSNP1370	0.054	1	Thecc.01G096000	Histone superfamily protein

Note. All F_{ST} values were significant at $P < 0.001$. Chr - chromosome. ¹ Pod index: number of fruits needed to produce 1 kg of dry seeds; a low pod index (e.g., 14-20) is preferred because it helps to select trees that yield heavier and larger seeds (Bekele et al. 2020). ² Gene ID and description based on Matina reference cacao genome (Theobroma cacao v2.1) (Motamayor et al. 2013) available at https://phytozome-next.jgi.doe.gov/info/Tcacao_v2_1

Three SNPs associated with yield traits showed high divergence between wild and cultivated cacao. Among these SNPs, Tcm004s00289192 ($F_{ST} = 0.160$, $P < 0.001$) was linked to dry bean weight and high yield, and TcSNP1370 ($F_{ST} = 0.054$, $P < 0.001$) was associated with number of seeds per fruit. Tcm002s23708704 ($F_{ST} = 0.107$, $P < 0.001$) was related to pod index, a measure of yield in commercial plantations. TcSNP1866 was associated with flower retention and had the highest $F_{ST} = 0.260$ ($P < 0.001$) between wild and managed cacao. Figure 4 shows the relative genotype frequency of SNPs showing pronounced divergence between wild and managed cacao and results of Fisher's exact test.

Managed and wild cacao showed also high divergence in disease resistance traits, especially in WBD resistance. We observed five SNPs significantly differentiated between managed and wild cacao that are associated with WBD resistance: Tcm006s19715703 ($F_{ST} = 0.146$, $P < 0.001$), Tcm004s00110232 ($F_{ST} = 0.101$, $P < 0.001$), TcSNP1230 ($F_{ST} = 0.083$, $P < 0.001$), Tcm003s33466269 ($F_{ST} = 0.067$, $P < 0.001$), and Tcm009s08066239 ($F_{ST} = 0.063$; $P < 0.001$). In addition, Tcm006s13222057 ($F_{ST} = 0.155$; $P < 0.001$) associated with resistance to CWC, and Tcm009s40465466 ($F_{ST} = 0.066$; $P < 0.001$) related to FPR resistance, were also significantly differentiated between managed and wild cacao (Table 3).

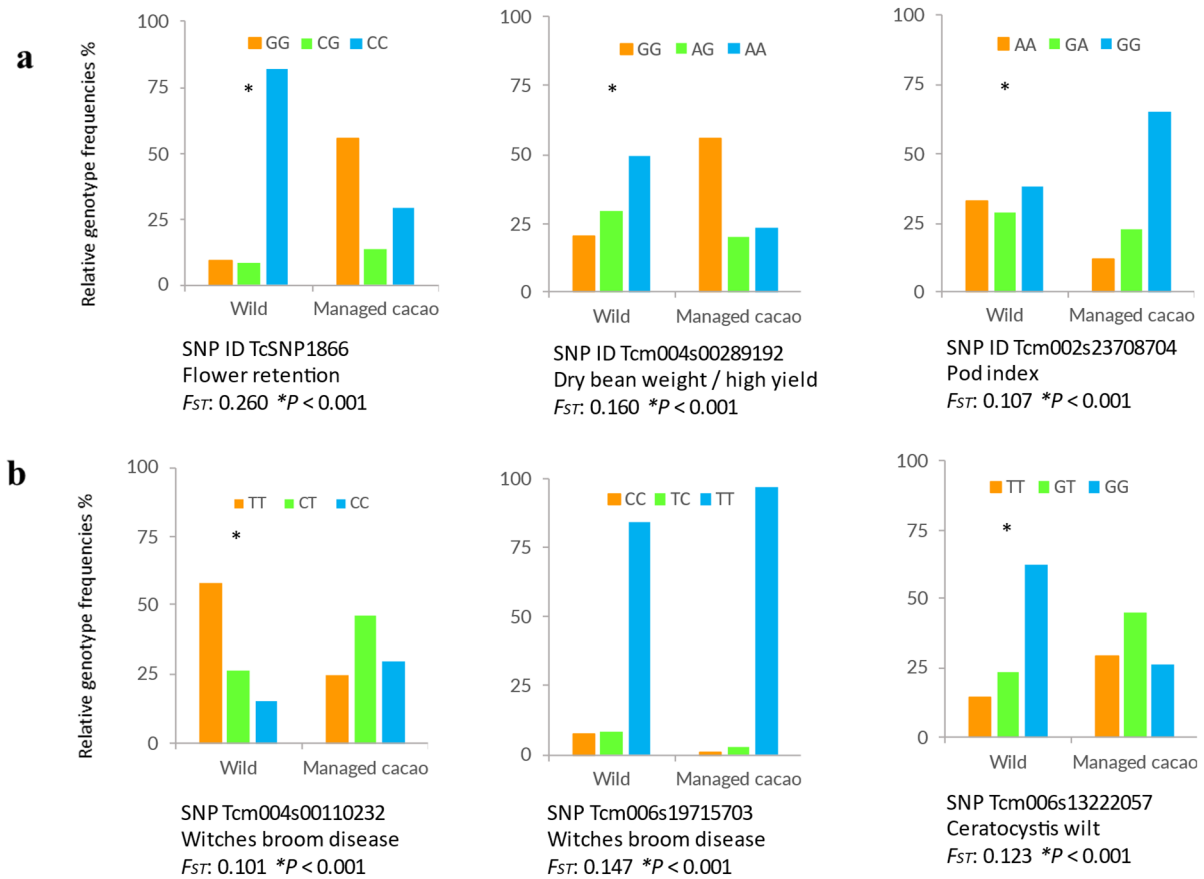


Fig. 4 Distribution of cacao genotypes for SNPs related to agronomic traits in the wild and managed cacao accessions. **a** - SNP markers related to yield traits: flower retention (da Silva et al. 2016); dry bean weight/high yield and high pod index (Fernandes et al. 2020). **b** - SNP markers related to disease resistance traits: Witches' broom disease (Royaert et al. 2016) and Ceratocystis wilt (Fernandes et al. 2018). *Significant difference between wild and managed cacao based on the Fisher's exact test ($P < 0.05$)

Figures 5 and 6 show the genotype frequencies for the SNPs Tcm004s00110232 and Tcm006s13222057 among cacao accessions resistant, tolerant and susceptible to WBD and CW, respectively. Germplasm with resistance ($n=58$) and tolerance ($n=25$) to WBD showed a significantly higher frequency of the CT and TT genotypes (Tcm004s00110232), and the CC genotype showed a low frequency in resistant plants, while all susceptible plants ($n=9$) showed the TT genotype (Figure 5). In addition, germplasm evaluated as resistant ($n=28$) and tolerant ($n=16$) to CW showed a higher frequency of GG and GT (Tcm006s13222057) genotypes, while the TT genotype was common in susceptible ($n=21$) plants. However, these differences were not statistically significant ($P > 0.05$, Figure 6).

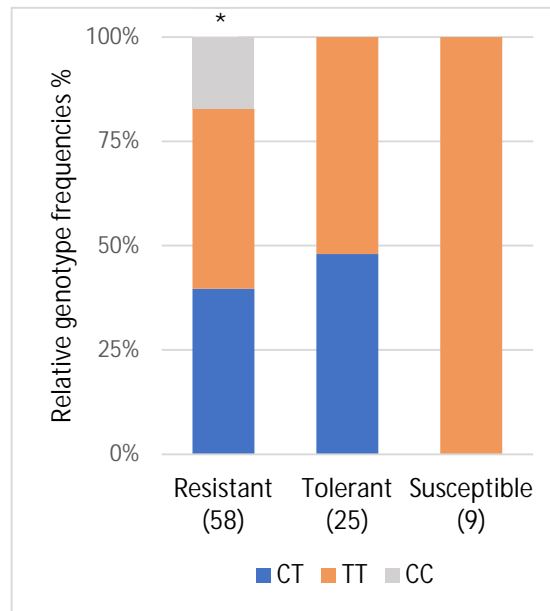


Fig. 5 Genotype frequencies for the SNP Tcm004s00110232 related to witches' broom disease (WBD) among cacao (*Theobroma cacao* L.) accessions with resistance, tolerance, or susceptibility to WBD. Number of accessions is provided in brackets. Reaction to WBD reported by the International Germplasm Cocoa Database (<https://www.icgd.reading.ac.uk/index.php>). *Significant difference between resistant-tolerant and susceptible based on the Fisher's exact test ($P = 0.0026$)

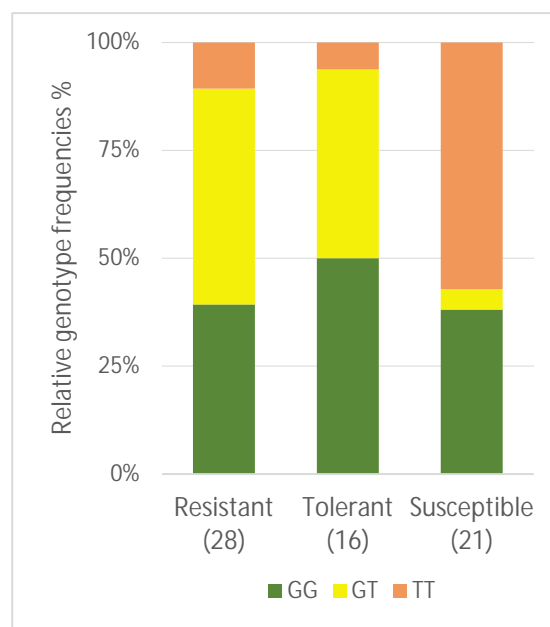


Fig. 6 Genotype frequencies for the SNP Tcm006s13222057 associated with Ceratosystis wild disease among cacao (*Theobroma cacao* L.) accessions with resistance, tolerance, or susceptibility to Ceratosystis disease. Number of accessions is provided in brackets. Reaction to Ceratosystis disease reported by the International Germplasm Cocoa Database (<https://www.icgd.reading.ac.uk/index.php>). No significant difference between resistant-tolerant and susceptible based on the Fisher's exact test ($P = 0.7910$)

4 DISCUSSION

Low population structure in managed cacao

With our set of SNPs, we did not identify the ten genetic clusters observed earlier by Motamayor et al. (2008) based on 96 simple sequence repeat (SSR) markers. It is not surprising, considering a limited set of nonrandom 42 SNPs that we used in our study, which could be also under selection. However, we still identified four clusters reflecting the geographic origin of wild samples, with the second and third clusters being overrepresented in managed cacao, mainly consisting of hybrids (Figure 1). This population pattern in managed cacao is in agreement with observations made by Cornejo et al. (2018). They analyzed the genome sequence of 200 cacao accessions and identified that cultivated cacao and man-made hybrids are mainly composed of two clusters, Criollo and Amelonado, with a low admixture of the Nacional cluster (Cornejo et al. 2018). Population genetic analysis demonstrated that Criollo was the first domesticated cacao and provides the foundations of cultivated cacao until today, mostly due to flavor and chocolate attributes (Lachenaud and Motamayor 2017; Motamayor et al. 2002). Our results confirm that a narrow cacao genetic diversity has been used in managed cacao likely due to retaining quality traits derived from Criollo observed in the second cluster in our analysis.

Cornejo et al. (2018) also observed that a higher Criollo ancestry in man-made hybrids is associated with low yield (seed productivity per year per plant) mainly due to the accumulation of deleterious mutations which led to reduced fitness in Criollo during the domestication process (Cornejo et al. 2018). The high contribution of Criollo to cultivated cacao helps to explain the low yield on average reported in cultivated cacao globally (Morrissey et al. 2019) and its susceptibility to BPR, WBD and FP diseases (Ploetz 2016). To capture new genetic diversity and to broaden the genetic base for future breeding activities, new plant collections

in northwestern Amazonia should be incorporated into cacao breeding programs (Nieves-Orduña et al. 2023).

In addition, the representation of the first and third clusters in managed cacao likely reflects gene introgression from the Pound Collection. The collection includes accessions MO, IMC, NA, SCA, PA collected in the Peruvian Amazon while searching for trees resistant to WBD (Zhang et al. 2011). The Pound collection has been widely used in cacao breeding for developing disease resistance after the collapse of plantations in Surinam, Trinidad, Ecuador and Brazil due to the introduction of WBD (Zhang et al. 2011). PA series also have been used for developing disease resistance against BPR and FPR (Zhang et al. 2011).

Genetic resources for cacao breeding

Guiana (GU) accessions were identified mainly in cluster four in our structure analysis and observed with a very low frequency in managed cacao (Figure 1). The underutilization of Guiana accessions in cultivated cacao was also observed by Cornejo et al. (2018). These findings highlight opportunities to exploit the GU accessions for cacao breeding. Early studies reported cacao trees from French Guiana as novel sources for resistance to BPR and WBD, and high yield (up to 1,426 kg of dry seeds/year/ha) (Lachenaud et al. 2007; Paulin et al. 2008). This was supported by Ofori et al. (2020), who observed that GU accessions can broaden the genetic base of cacao breeding not only for BPR resistance but also yield in Ghana (Ofori et al. 2020). In addition, Guiana accessions in Central America showed moderate resistance to FPR (Lachenaud et al. 2018), an essential agronomic trait for cacao cultivation in Tropical America (Evans et al. 1977; Gutiérrez et al. 2021b; Phillips-Mora et al. 2005, 2013). A detailed agronomic evaluation of multiple Guiana accessions is presented by Lachenaud et al. (2007). The best clones and those to be avoided were identified based on yield, disease resistance (BPR and WBD) and seed quality traits. The preselection of the best Guiana germplasm facilitates

introgression of new and valuable genetic diversity resources into cacao breeding programs (Lachenaud et al. 2007).

Yield and disease resistance: agronomic traits showing divergence

Cacao yield traits and disease resistance to WBD showed patterns of divergence between wild and managed cacao accessions (Table 3). These patterns are associated with the history of cacao domestication, cultivation and selection. After initial selection for pulp flavor and seed traits in the sister Curaray population on the Ecuador-Colombia border, human selection led to the domestication of Criollo cacao ~3600 years ago (Clement et al. 2010; Cornejo et al. 2018). In addition, evidence based on ancient cacao DNA supports the consumption of cacao in the Ecuadorian Amazon around 5300 years ago, the DNA analyzed was closer to the Curaray and Purus clusters than to other cacao clusters. (Zarrillo et al. 2018). The genetic cost of cacao domestication led to the accumulation of deleterious mutations, susceptibility to diseases and low yield in Criollo (Cornejo et al. 2018). Criollo with larger seed size, white cotyledons and reduced bitterness was distributed and cultivated outside the Upper Amazon by native Americans in Northern Colombia and Mesoamerica (Cornejo et al. 2018; Motamayor et al. 2002).

Criollo cultivation expanded during colonial times through Tropical America (Bartley 2005), especially in Trinidad, where some blast destroyed the crop in 1727 (Díaz-Valderrama et al. 2020). This collapse in cacao production led to introduction of new plant material from upper Amazonia, which hybridized naturally with the cultivated Criollo and formed a hybrid cultivated cacao known as Trinitario germplasm (Zhang et al. 2011). From these vigorous hybrids, a breeding program was started in 1930 by the Imperial College of Tropical Agriculture of Trinidad focused on yield traits (Toxopeus 1969). The best trees were selected based on the number of seeds per pod and bean weight, which resulted in 100 trees known as the Imperial College Selection (Toxopeus 1969). The mean seed weight was an important trait

due to the premium price for large seeds in the market (Toxopeus 1969), a trait that still is considered of economic importance in modern cacao breeding programs (Bekele et al. 2022). In addition, after the introduction and impact of WBD in 1932, new wild material with disease-resistance traits was needed in Trinidad. The material searched and collected from the Peruvian Amazon in 1937-1938 includes accessions known as the Pound Collection. This collection created the genetic base for developing breeding resistance against WBD globally (Bartley 2005; Díaz-Valderrama et al. 2020; Evans 2016; Zhang et al. 2011).

In Brazil, the low genetic diversity of cultivated cacao led to the collapse of cacao economies in 1989 due to introduction of WBD in Bahia (Bennett 2003; Evans 2016). In response, breeding programs started in Brazil to broaden the genetic base of cultivated cacao and evaluated germplasm collections for developing resistance against WBD (Bennett 2003).

SNPs related to flower setting (sexual compatibility)

Within accessions analyzed here we observed TcSNP1866 with the highest F_{ST} value of 0.260 showing pronounced divergence between wild and managed cacao (Table 3). This SNP was identified in a GWAS study using 295 trees and 5301 SNPs, and incompatibility was measured as frequency of flower retention 15 days after self-pollination (on average 21 flowers were self-pollinated per tree) instead of a yes/no trait (da Silva et al. 2016). Previous studies in cacao highlighted that thousands of individuals are necessary to avoid bias in the estimation of SNP effects associated with small samples sizes (da Silva et al. 2016). In addition, flower dropping is influenced by rainfall, high temperature or insects' attack which can introduce underestimation in flower retention values (da Silva et al. 2016). Thus, the effect of genotype CC (TcSNP1866) associated with a high percentage (33%) of flower retention may be influenced by small sampling sizes and/or environmentally induced flower dropping (da Silva et al. 2016). From the breeders' perspective, to select self-compatible trees it is recommended to implement genomic selection which considers thousands of SNPs (da Silva et al. 2016).

Clones common to cacao breeding, including disease resistant and commercial trees are self-incompatible (López et al. 2021; Phillips-Mora et al. 2013). Self-incompatibility in cacao requires plantation designs where cross compatible clones are established to foster the exchange of pollen and field production (López et al. 2021; Phillips-Mora et al. 2013). Since incompatibility is a limiting factor in cacao yield, a common breeding objective is to avoid self-incompatible trees in breeding populations (López et al. 2021).

Although we do not have information on the percentage of flower retention after self-pollinations in the accessions genotyped here that would allow us to validate the phenotypes associated with TcSNP1866, we observed a significantly lower frequency of the genotype CC associated with self-compatibility (da Silva et al. 2016) in managed cacao (Figure 4A), which may reflect the fact that breeding populations and advanced selections include self-incompatible trees. For example, self-compatibility was evaluated in commercial clones such as EET (62, 95, 96, 400), CAUCASIA (37, 39, 43, 47), ICS (1,6,39,60,95), and UF (29, 273, 613, 667, 676) (López et al. 2021). These clones (5 to 10 years old) were self-pollinated, and the mean fruit set was 24% across clones, indicating partial self-incompatibility (López et al. 2021). The genotypes (n=18) for these incompatible clones indicated a higher proportion of the homozygous GG (56%), followed by CC (39%) and CG (6%).

As self-compatibility is not absolute in cacao (Lopes et al. 2022), the observation of the genotype CC (TcSNP1866) at higher frequency in wild cacao reflect some degree of self-compatibility in wild populations (Figure 4A). This could be due to geographic origin of wild trees observed isolated along river basins and pollinated by midges with reduced range of movement (Lopes et al. 2022). Accordingly, levels of homozygosity above 70% were observed among wild accessions genotyped with genome-wide SNPs (3K) (Lopes et al. 2022).

SNPs related to yield and pod index

Cacao yield (kg of dry seeds/year/ha) is a polygenic complex trait, at least 40 candidate genes of functional importance encode embryo and seed development, protein synthesis, carbohydrate transport, and lipid biosynthesis and transport (Bekele et al. 2022). In addition, yield components such as number of pods produced per tree, and bean dry weight per pod are influenced by genotypes, environment, agronomic management (e.g., fertilization, shade, irrigation) and diseases (Bartley 2005; Doaré et al. 2020; Fernandes et al. 2020; Phillips-Mora et al. 2013; Solís Bonilla et al. 2022). Disease pressure also influences yield, as high-yielding clones with disease resistance traits exhibited a low diseases incidence (Phillips-Mora et al. 2013). But the same clone can exhibit varying yield potential due to exposition to different pathogen strains (variations in disease pressure) (Jaimes et al. 2011; Jaimes et al. 2019). Disease management in cacao farms also affects the yield potential of clones (Jaimez et al. 2020). For example, removal of FPR-infected pods is a common practice to reduce disease incidence and increase tree productivity (Jaimes et al. 2019; Jaimez et al. 2020).

The SNP Tcm004s00289192 related to increased high yield showed a pronounced divergence between wild and managed cacao (Table 3). Within accessions analyzed here we observed the homozygous genotype GG more frequently in managed cacao (Figure 4A). Among the accessions with the GG genotype are advanced selections such as ICS, UF, TARS and EET clones. The heterozygous genotype AG was observed in high-yielding clones such as TARS, CATIE R6, CCN 51 and VB clones. The allele G (Tcm004s00289192) is associated with a higher yield (Fernandes et al. 2020). Fernandes et al. (2020) identified copy-number variations of SWEET (Sugar Will Eventually be Exported Transporters) genes between the markers Tcm004s00289192 and Tcm004s00615809 on chromosome IV. Likewise, a recent GWAS using a diverse cacao germplasm in Trinidad identified SNPs markers within the genes SWEET17 (chromosome 4) and SWEET2 (chromosome 7) related to cacao yield traits such as

pod index and seed number (Bekele et al. 2022). Sweet proteins are a family of sugar transporters important for plants biological processes such as growth, development, and response to abiotic and biotic stresses (Singh et al. 2023). SWEET genes in cultivated tree species such as apple (*Malus x domestica*) contribute to fruit sugar accumulation (Zhen et al. 2018), and in *Litchi chinensis* played roles in fruit development, growth and seed development (Xie et al. 2019). SWEET4 also contributed to rice and maize domestication by enhancing seed sizes as it facilitates sugar transport during grain filling (Sosso et al. 2015). Further characterization of cacao SWEET genes in wild and cultivated cacao can help to identify alleles contributing to cacao domestication and validate useful alleles for improving yield traits in cacao.

In cacao breeding pod index is defined as the number of pods (fruits) necessary to produce one kg of dry seeds, a low pod index (14-20) is associated with high yield potential and heavier seeds (Bekele et al. 2020; Fernandes et al. 2020). Early breeding programs in Trinidad (1940s) focused on yield and pod index; selected trees exhibited a low pod index (18) and higher yield (1000kg/ha) (Toxopeus 1969). The pod index also showed a high narrow-sense heritability (0.64) and stability across different sites at two farms in Bahia, Brazil, making the trait a target of selection (DuVal et al. 2017). In addition, a low pod index is also selected in germplasm collections and breeding populations because less healthy fruits are necessary to produce one kg of dry cacao, meaning less costs associated with harvesting and pod breaking (Bekele et al. 2020; Solís Bonilla et al. 2022). The SNP Tcm002s23708704 related to pod index showed divergence between wild and managed cacao (Table 3), and Fernandes et al. (2020) reported the A allele to be associated with high pod index. We observed the genotype AA (SNP Tcm002s23708704) at low frequency in managed cacao compared to wild cacao (Figure 4A). Although we do not have phenotypic information for the accession genotyped here, the low

frequency of the A allele in managed cacao may reflect the continued selection favoring a low pod index in breeding populations and cultivated cacao.

SNPs related to witches' broom disease (WBD) resistance

WBD is a devastating disease for cacao cultivation (Evans 2016). To improve disease resistance, cacao breeders use SCA 6 and SCA 12 clones as primary sources for breeding WBD resistance (Gutiérrez et al. 2016). However, due to the susceptibility of SCA clones under high disease pressure, there was a need to identify new sources of resistance. In this effort, Pereira et al. (2021) identified new clones, such as C SUL-3 and GU-171, resistant to WBD in Bahia, Brazil. In addition, new expeditions to the Peruvian Amazon collected 280 cacao trees to diversify the gene pool and resistance to WBD (Durham 2011). Native to Amazonia but widely distributed in South America, WBD has different pathogen strains (Lisboa et al. 2020; Ploetz 2016). Thus, a major breeding objective is to develop cultivars with broad WBD disease resistance (Meinhardt et al. 2008). WBD resistance is a complex trait involving at least sixteen candidate resistance genes identified using genome-wide association studies (GWAS) (Osorio-Guarín et al. 2020) and quantitative trait locus (QTL) analysis (Chia Wong et al. 2022; Mournet et al. 2020; Royaert et al. 2016).

The SNP Tcm004s00110232 related to WBD resistance showed divergence between wild and managed cacao (Table 3). We inferred the data on reaction to WBD for 92 samples (30%) from the ICGD (Turnbull and Hadley 2023), while 219 (70%) have no data available. The phenotypic data reported are from different WBD studies (Turnbull and Hadley 2023), and due to region specific differences among studies likely responses to different pathogen strains were assessed. The genotypes CT and TT (Tcm004s00110232) are highly represented in resistant (n=58) and tolerant (n=25) plants (Figure 5). The genotype CT observed in the clone TSH 1188 was reported as WBD resistant by Royaert et al. (2016). In addition, the genotype CC (Tcm004s00110232) was observed only in resistant germplasm (Figure 5). These results are in

agreement with Lachenaud et al. (2007) that reported the observed homozygous CC in resistant to WBD Guiana clones (GU 171 /C; GU 219 /F; GU 221 /C; GU 261 /P; GU 277 /G) (Lachenaud et al. 2007; Turnbull and Hadley 2023).

The genotype TT (Tcm004s00110232) was detected in all nine accessions reported as susceptible to WBD (Figure 5) but with more samples for this category with other genotypes could probably be detected. This reaction to WBD is explained by distinct variants of *M. pernicioso* (range of pathogen aggressiveness) and pathogen adaptation to resistance trees such as SCA 6 (Artero et al. 2017; Pereira et al. 2021; Royaert et al. 2016). For example, in Bahia, Brazil, a decrease in resistance of SCA clones to WBD has been reported due to continuous and high disease pressure (Pereira et al. 2021). The fact that the SCA clones have been the main source for breeding WBD resistance motivated the search and evaluation of new sources of disease resistance from different geographical areas (Durham 2011; Pereira et al. 2021). In addition, disease resistance is a polygenic trait with each gene explaining a relatively small portion of the variation in disease resistance.

The majority of wild accessions here genotyped correspond to a subset of samples collected during the 1930-1940s in Peru by Frederick J. Pound (see material and methods), whose main purpose was collecting cacao trees exhibiting WBD disease resistance traits (Bartley 2005; Zhang et al. 2011). This explains why the favorable allele T (Tcm004s00110232) was observed at high frequency in wild cacao (Figure 4B). The high frequency of genotype CT in managed cacao reflect the hybridization of managed cacao with wild cacao accessions. For example, a total of 191 crosses used SCA 6 (low yield but resistant to WBD) as a parent for the incorporation of disease resistance traits in cultivated cacao (Turnbull and Hadley 2023). As a result, commercial hybrids such as ICS, TSH, EET, EQX, and TARS with various levels of WBD resistance were developed using SCA 6 (Turnbull and Hadley 2023).

SNPs related to *Ceratocystis* wilt (CW) resistance

CW targets the cacao vascular system and causes the death of infected trees (Engelbrecht et al. 2007). The disease is caused by the host-specialized fungus *Ceratocystis cacaofunesta*, which is native to South America (Western Ecuador and Southwest Brazil) (Engelbrecht et al. 2007). The disease is geographically restricted to Tropical America. Still, it threatens the cacao economy because it can be dispersed to important cacao-producing regions such as West Africa and Asia (Engelbrecht et al. 2007). Early reports in the 1950s described CW causing damage to cocoa farms in Colombia, Ecuador, Costa Rica and Trinidad, and in 1997 it was observed in Bahia, Brazil (Cabrera et al. 2016). Breeding for disease resistance is gaining attention as germplasm selected for WBD resistance in Brazil such as “Theobahia” shows susceptibility to CW (Fernandes et al. 2018; Lopes et al. 2011).

The SNP Tcm006s13222057 is associated with resistance to CW (Fernandes et al. 2018). This SNP shows significant differences between wild and cultivated cacao populations (Table 3). To investigate the relationship between SNP genotypes and CW resistance, we searched in the ICGD for the reaction to CW among the accessions genotyped (Turnbull and Hadley 2023). Only 65 accessions (21%) have reported data on the CW reaction (resistant, tolerant or susceptible), while 249 accessions (79%) have no data available. Among the accessions with reported data, we observed that the resistant (n=28) and tolerant (n=16) groups had a high frequency of the G allele (Tcm006s13222057), being the genotypes GT and GG more frequent in the resistant and tolerant group, respectively (Figure 6). While in the susceptible group the genotype TT is prevalent (Figure 6). However, these differences were not significant, likely, due to small sample size. These results are however consistent with the QTL analysis by Fernandes et al. (2018), which identified the G allele as a marker for CW resistance.

We observed the genotype GG at high frequency in wild cacao (Figure 4B). Among these accessions the IMC clones (IMC 11, 31, 47, 60 and 67), PA 121, POUND (12,12A), SCA (6,

12) and U (26, 70) are reported as resistant to CW (Turnbull and Hadley 2023). The wild accession IMC 67 is frequently used in cacao breeding programs for its resistance to CW and vegetative vigor and is widely used as a rootstock (Cabrera et al. 2016; Osorio Montoya et al. 2022). IMC 67 was used in 145 crosses for developing cultivars and breeding populations of the series CEPEC, EET, EQX, TSH, TSH (Turnbull and Hadley 2023). However, IMC 67 was reported to be resistant to the Ecuadorian isolate but susceptible to the Brazilian strain (Cabrera et al. 2016), highlighting the need for new (wild) sources against different CW strains. In addition, we observed the heterozygous genotype GT at high frequency in managed cacao (Figure 4B). Heterozygous clones reported as disease resistant to CW includes EET clones (399, 400), SC 20, ICS clones (6, 40, 95), TSH (1188, 595), UF (29, 650) and VB 650 and VB 681, which is recommended for large scale plating in Bahia, Brazil (Lopes et al. 2011; Turnbull and Hadley 2023).

5 CONCLUSIONS AND RESEARCH QUESTIONS

Accessions structure of wild and managed cacao

Our results confirm a narrow genetic diversity in managed cacao likely reflecting Criollo ancestry of accessions traditionally used to select for chocolate quality. Among the germplasm analyzed, managed cacao (cultivated and breeding populations) showed introgression of wild cacao collected in western Amazonia (Peru, Ecuador), but much less contribution of Guiana clones. New collection trips should be done to broaden the genetic base for cultivated cacao to improve agronomic traits, such as urgently needed FPR disease resistance. Potential areas for cacao germplasm collection are highlighted by Nieves et al. (2023).

New genetic resources for cacao breeding

Guiana accessions represent a genetic resource underutilized in managed cacao. Agronomic evaluations of this germplasm group showed favorable yield traits (e.g., accession GU 285)

and resistance against WBD and BPR (Lachenaud et al. 2007). Cacao breeding programs can obtain Guiana clones free of pathogens through the ICQC and use them in perspective breeding experiments to exploit their potential under a hybridization scheme. Developing of cacao inbred lines can support hybrid breeding but homozygosity and genetic distance between accessions should be considered to maximize heterosis (Akpertey et al. 2022; Lopes et al. 2022).

Validation of SNPs and cacao breeding

Cacao SNPs associated with disease resistance and yield traits should be validated across diverse germplasm added with standard phenotypic information obtained from multi-environment studies and using trees with the same age as well as clonal replicates. This validation facilitates the identification of useful major QTL for cultivar development by different breeding programs. In addition, breeding programs can adopt and optimize genomic selection. Previous studies have demonstrated to improve selection if individual effects of multiple SNP markers are considered in complex cacao traits such as yield and disease resistance (Bekele et al. 2022; McElroy et al. 2018; Romero-Navarro et al. 2017).

Candidate genes of cacao domestication

Disease resistance and yield traits showed divergence between wild and managed cacao, probably reflecting selection during domestication, cultivation and breeding efforts. Further analysis with a diverse natural cacao population, cultivars (e.g., CEPEC's germplasm of Brazil) and clones developed by chocolate companies for specific farm conditions, using high density SNP genotyping panels (cacao 15K SNP array) will help identify signatures of selection in candidate genes with high phenotypic impact on agronomic traits.

Other traits of economic importance

Plant architecture is a trait of economic importance not widely investigated in breeding programs to improve harvest index (ratio of seeds to total below biomass produced). The breeding challenge is to reduce plant height to facilitate harvesting but without affecting yield. Mustiga et al. (2018) proposed selecting trees with small trunk diameter, lower branch angles, and high yield to increase plant density. In addition, market restrictions in cacao-based products with high cadmium levels have led to evaluating germplasm and identifying valuable genetic sources to introduce low cadmium traits into commercial clones (Lewis et al. 2018).

Supplementary Information The online version contains supplementary material available at

Author contributions HEN-O participated in conceptualization, funding acquisition, investigation, data analysis, figure elaboration, writing original draft, validation, reviewing and editing the manuscript; MM participated in data analysis, validation, reviewing and editing the manuscript; KVK participated in conceptualization, investigation, data analysis, supervision, validation, reviewing and editing the manuscript; OG participated in conceptualization, funding acquisition, investigation, data analysis, project administration, supervision, validation, reviewing and editing the manuscript. All authors commented and edited intermediate versions of the manuscript. All authors reviewed and approved the final manuscript.

Funding We thank the German Academic Exchange Service-DAAD for the financial support provided to Helmuth E. Nieves-Orduña. Open access funding was enabled and organized by the Project DEAL.

Data availability This research used only published SNP markers, which are summarized in table 1. The genotype data generated within this study is presented in supplementary table S3.

Acknowledgements The authors would like to thank the International Center for Tropical Agriculture (CATIE) and The International Cocoa Quarantine Centre at the Reading (ICQC,

R) for providing the cacao leaf samples for DNA analysis. We acknowledge support by the Open Access Publication Funds of the Georg-August University of Göttingen. We thank anonymous reviewers for insightful suggestions, which helped us improve the manuscript, and editors for handling the manuscript.

Declarations

Conflict of interest The authors declare no competing interests

Data Archiving Statement This research used only published SNP markers, which are summarized in table 1. The genotype data generated within this study is presented in supplementary table S3.

6 REFERENCES

- Agena Bioscience (2015) Assay Design Suite v2.0: User Guide. USG-CUS-040 Rev02
- Agena Bioscience (2019a) iPLEX Pro and Gold Reagents: User Guide. USG-CUS-071 Rev03
- Agena Bioscience (2019b) MassARRAY Typer: User Guide. USG-CUS-027 Rev05
- Akperterey A, Padi FK, Meinhardt L, Zhang D (2022) Relationship between genetic distance based on single nucleotide polymorphism markers and hybrid performance in Robusta coffee (*Coffea canephora*). *Plant Breeding* 141:286–300. <https://doi.org/10.1111/pbr.13005>
- Allen JB (1988) Geographical variation and population biology in wild *Theobroma cacao*. PhD, The University of Edinburgh
- Artero AS, Silva JQ, Albuquerque PSB, Bressan EA, Leal GA, Sebbenn AM, Griffith GW, Figueira A (2017) Spatial genetic structure and dispersal of the cacao pathogen *Moniliophthora perniciosa* in the Brazilian Amazon. *Plant Pathol* 66:912–923. <https://doi.org/10.1111/ppa.12644>
- Barlow J, Gardner TA, Lees AC, Parry L, Peres CA (2012) How pristine are tropical forests? An ecological perspective on the pre-Columbian human footprint in Amazonia and implications for contemporary conservation. *Biological Conservation* 151:45–49. <https://doi.org/10.1016/j.biocon.2011.10.013>
- Bartley BGD (2005) The genetic diversity of cacao and its utilization. CABI Pub, Cambridge MA
- Bekele F, Phillips-Mora W (2019) Cacao (*Theobroma cacao* L.) breeding. In: Al-Khayri JM, Jain SM, Johnson DV (eds) *Advances in Plant Breeding Strategies: Industrial and Food Crops*, vol 46. Springer International Publishing, Cham, pp 409–487
- Bekele F, Bidaisee G, Singh H, Saravanakumar D (2020) Morphological characterisation and evaluation of cacao (*Theobroma cacao* L.) in Trinidad to facilitate utilisation of Trinitario cacao globally. *Genet Resour Crop Evol* 67:621–643. <https://doi.org/10.1007/s10722-019-00793-7>
- Bekele F, Bidaisee G, Allegre M, Argout X, Fouet O, Boccara M, Saravanakumar D, Bekele I, Lanaud C (2022) Genome-wide association studies and genomic selection assays made in a large sample of cacao (*Theobroma cacao* L.) germplasm reveal significant marker-trait associations and good predictive value for improving yield potential. *PLoS ONE* 17:e0260907
- Bennett AB (2003) Out of the Amazon: *Theobroma cacao* enters the genomic era. *Trends Plant Sci* 8:561–563. <https://doi.org/10.1016/j.tplants.2003.10.004>
- Boza EJ, Motamayor JC, Amores FM, Cedeño-Amador S, Tondo CL, Livingstone DS, Schnell RJ, Gutiérrez OA (2014) Genetic characterization of the cacao cultivar CCN 51: its impact and significance on global cacao improvement and production. *Journal of the American Society for Horticultural Science* 139:219–229. <https://doi.org/10.21273/JASHS.139.2.219>
- Cabrera OG, Molano EPL, José J, Álvarez JC, Pereira GAG (2016) Ceratocystis Wilt pathogens: history and biology—highlighting *C. cacaofunesta*, the causal agent of Wilt Disease of cacao. In: Bailey BA, Meinhardt LW (eds) *Cacao Diseases*. Springer International Publishing, Cham, pp 383–428
- Chia Wong JA, Clement DPL, Mournet P, dos Santos Nascimento A, Solís Bonilla JL, Lopes UV, Pires JL, Gramacho KP (2022) A high-density genetic map from a cacao F2 progeny and QTL detection for resistance to witches' broom disease. *Tree Genetics & Genomes* 18. <https://doi.org/10.1007/s11295-022-01562-x>

- Clement C, Cristo-Araújo M de, Coppens D'Eeckenbrugge G, Alves Pereira A, Picanço-Rodrigues D (2010) Origin and domestication of native Amazonian crops. *Diversity* 2:72–106. <https://doi.org/10.3390/d2010072>
- Clement CR, Denevan WM, Heckenberger MJ, Junqueira AB, Neves EG, Teixeira WG, Woods WI (2015) The domestication of Amazonia before European conquest. *Proc. R. Soc. Lond. B* 282:20150813. <https://doi.org/10.1098/rspb.2015.0813>.
- Coe SD, Coe MD (2013) *The true history of chocolate*. Thames & Hudson, London
- Cornejo OE, Yee M-C, Dominguez V, Andrews M, Sockell A, Strandberg E, Livingstone D, Stack C, Romero A, Umaharan P, Royaert S, Tawari NR, Ng P, Gutierrez O, Phillips W, Mockaitis K, Bustamante CD, Motamayor JC (2018) Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process. *Commun Biol* 1:167. <https://doi.org/10.1038/s42003-018-0168-6>
- da Silva MR, Clément D, Gramacho KP, Monteiro WR, Argout X, Lanaud C, Lopes U (2016) Genome-wide association mapping of sexual incompatibility genes in cacao (*Theobroma cacao* L.). *Tree Genetics & Genomes* 12:15. <https://doi.org/10.1007/s11295-016-1012-0>
- Dand R (2011) *The international cocoa trade*. Elsevier, Cambridge UK
- Daymond A (2018) Safe handling and movement of cocoa germplasm for breeding. In: Umaharan P (ed) *Achieving sustainable cultivation of cocoa*. Burleigh Dodds Science Publishing, Cambridge UK, pp 101–110
- Díaz-Valderrama JR, Leiva-Espinoza ST, Aime MC (2020) The history of cacao and its diseases in the Americas. *Phytopathology* 110:1604–1619. <https://doi.org/10.1094/PHYTO-05-20-0178-RVW>
- Doaré F, Ribeyre F, Cilas C (2020) Genetic and environmental links between traits of cocoa beans and pods clarify the phenotyping processes to be implemented. *Sci Rep* 10:9888. <https://doi.org/10.1038/s41598-020-66969-9>
- Durham S (2011) Peruvian cacao collections trip yields treasures. *Agricultural Research* 59:8–10
- DuVal A, Gezan SA, Mustiga G, Stack C, Marelli J-P, Chaparro J, Livingstone D, Royaert S, Motamayor JC (2017) Genetic parameters and the impact of off-types for *Theobroma cacao* L. in a breeding program in Brazil. *Front Plant Sci* 8:2059. <https://doi.org/10.3389/fpls.2017.02059>
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet Resour* 4:359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Engelbrecht CJ, Harrington TC, Alfenas A (2007) Ceratocystis wilt of cacao—a disease of increasing importance. *Phytopathology* 97:1648–1649. <https://doi.org/10.1094/PHYTO-97-12-1648>
- Evans HC, Edwards DF, Rodríguez M (1977) Research on Cocoa Diseases in Ecuador: Past and Present. *PANS* 23:68–80. <https://doi.org/10.1080/09670877709412399>
- Evans HC (2016) Witches' Broom Disease (*Moniliophthora perniciosa*): History and Biology. In: Bailey BA, Meinhardt LW (eds) *Cacao Diseases*. Springer International Publishing, Cham, pp 137–177
- Fernandes LDS, Royaert S, Corrêa FM, Mustiga GM, Marelli J-P, Corrêa RX, Motamayor JC (2018) Mapping of a major QTL for *Ceratocystis* Wilt disease in an F1 population of *Theobroma cacao*. *Front Plant Sci* 9:155. <https://doi.org/10.3389/fpls.2018.00155>
- Fernandes LDS, Correa FM, Ingram KT, Almeida A-AFd, Royaert S (2020) QTL mapping and identification of SNP-haplotypes affecting yield components of *Theobroma cacao* L. *Hortic Res* 7:26. <https://doi.org/10.1038/s41438-020-0250-3>

- Gutiérrez O, Campbell A, Phillips-Mora W (2016) Breeding for disease resistance in cacao. In: Bailey BA, Meinhardt LW (eds) Cacao Diseases. Springer International Publishing, Cham, pp 567–609
- Gutiérrez O, Martínez K, Zhang D, Livingstone D, Turnbull CJ, Motamayor JC (2021a) Selecting SNP markers reflecting population origin for cacao (*Theobroma cacao* L.) germplasm identification. Beverage Plant Research 1:1–9. <https://doi.org/10.48130/BPR-2021-0015>
- Gutiérrez O, Puig A, Phillips-Mora W, Bailey BA, Ali SS, Mockaitis K, Schnell RJ, Livingstone D, Mustiga G, Royaert S, Motamayor JC (2021b) SNP markers associated with resistance to frosty pod and black pod rot diseases in an F1 population of *Theobroma cacao* L. Tree Genetics & Genomes 17:106. <https://doi.org/10.1007/s11295-021-01507-w>
- Hoang NT, Kanemoto K (2021) Mapping the deforestation footprint of nations reveals growing threat to tropical forests. Nat Ecol Evol 5:845–853. <https://doi.org/10.1038/s41559-021-01417-z>
- Jaimes Y, Aránzazu F, Rodríguez E, Martínez N (2011) Behavior of introduced regional clones of *Theobroma cacao* toward the infection *Moniliophthora roreri* in three different regions of Colombia. Agron. colomb. 29:361–371
- Jaimes Y, Ribeyre F, Gonzalez C, Rojas J, Furtado E, Cilas C (2019) Factors affecting the dynamics of Frosty Pod Rot in the main cocoa areas of Santander State, Colombia. Plant Dis 103:1665–1673. <https://doi.org/10.1094/PDIS-10-18-1761-RE>
- Jaimez R, Vera D, Mora A, Loor R, Bailey B (2020) A disease and production index (DPI) for selection of cacao (*Theobroma cacao*) clones highly productive and tolerant to pod rot diseases. Plant Pathology 69:698–712. <https://doi.org/10.1111/ppa.13156>
- Kalischek N, Lang N, Renier C, Daudt RC, Addoah T, Thompson W, Blaser-Hart WJ, Garrett R, Schindler K, Wegner JD (2023) Cocoa plantations are associated with deforestation in Côte d’Ivoire and Ghana. Nat Food 4:384–393. <https://doi.org/10.1038/s43016-023-00751-8>
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Lachenaud P (1997) Genetic/taxonomic structuring of the *Theobroma cacao* L. species fresh hypotheses. INGENIC Newsletter 3:10–11
- Lachenaud P, Paulin D, Ducamp M, Thevenin J-M (2007) Twenty years of agronomic evaluation of wild cocoa trees (*Theobroma cacao* L.) from French Guiana. Scientia Horticulturae 113:313–321. <https://doi.org/10.1016/j.scienta.2007.05.016>
- Lachenaud P (2015) Wild cacao trees (*Theobroma cacao* L.) in French Guiana. Academic Journal of Suriname 6:555–560
- Lachenaud P, Motamayor JC (2017) The Criollo cacao tree (*Theobroma cacao* L.): a review. Genet Resour Crop Evol 64:1807–1820. <https://doi.org/10.1007/s10722-017-0563-8>
- Lachenaud P, Clément D, Argout X, Scalabrin S, Doaré F (2018) The Guiana cacao genetic group (*Theobroma cacao* L.): a new core collection in French Guiana. Botany Letters 165:248–254. <https://doi.org/10.1080/23818107.2018.1465466>
- Langella O (2001) Populations. Free Software Foundation. <https://bioinformatics.org/populations/>
- Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>
- Levis C, Costa FRC, Bongers F, Peña-Claros M, Clement CR, Junqueira AB, Neves EG, Tamanaha EK, Figueiredo FOG, Salomão RP, Castilho CV, Magnusson WE, Phillips OL, Guevara JE, Sabatier D, Molino J-F, López DC, Mendoza AM, Pitman NCA, Duque A,

- Vargas PN, Zartman CE, Vasquez R, Andrade A, Camargo JL, Feldpausch TR, Laurance SGW, Laurance WF, Killeen TJ, Nascimento HEM, Montero JC, Mostacedo B, Amaral IL, Guimarães Vieira IC, Brienen R, Castellanos H, Terborgh J, Carim MdJV, Guimarães JRdS, Coelho LdS, Matos FDdA, Wittmann F, Mogollón HF, Damasco G, Dávila N, García-Villacorta R, Coronado ENH, Emilio T, Filho DdAL, Schiatti J, Souza P, Targhetta N, Comiskey JA, Marimon BS, Marimon B-H, Neill D, Alonso A, Arroyo L, Carvalho FA, Souza FC de, Dallmeier F, Pansonato MP, Duivenvoorden JF, Fine PVA, Stevenson PR, Araujo-Murakami A, Aymard C GA, Baraloto C, do Amaral DD, Engel J, Henkel TW, Maas P, Petronelli P, Revilla JDC, Stropp J, Daly D, Gribel R, Paredes MR, Silveira M, Thomas-Caesar R, Baker TR, da Silva NF, Ferreira LV, Peres CA, Silman MR, Cerón C, Valverde FC, Di Fiore A, Jimenez EM, Mora MCP, Toledo M, Barbosa EM, Bonates LCdM, Arboleda NC, Farias EdS, Fuentes A, Guillaumet J-L, Jørgensen PM, Malhi Y, Andrade Miranda IP de, Phillips JF, Prieto A, Rudas A, Ruschel AR, Silva N, Hildebrand P von, Vos VA, Zent EL, Zent S, Cintra BBL, Nascimento MT, Oliveira AA, Ramirez-Angulo H, Ramos JF, Rivas G, Schöngart J, Sierra R, Tirado M, van der Heijden G, Torre EV, Wang O, Young KR, Baider C, Cano A, Farfan-Rios W, Ferreira C, Hoffman B, Mendoza C, Mesones I, Torres-Lezama A, Medina MNU, van Andel TR, Villarreal D, Zagt R, Alexiades MN, Balslev H, Garcia-Cabrera K, Gonzales T, Hernandez L, Huamantupa-Chuquimaco I, Manzatto AG, Milliken W, Cuenca WP, Pansini S, Pauletto D, Arevalo FR, Reis NFC, Sampaio AF, Giraldo LEU, Sandoval EHV, Gamarra LV, Vela CIA, Steege H (2017) Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science* 355:925–931. <https://doi.org/10.1126/science.aal0157>
- Lewis C, Lennon AM, Eudoxie G, Umaharan P (2018) Genetic variation in bioaccumulation and partitioning of cadmium in *Theobroma cacao* L. *Sci Total Environ* 640-641:696–703. <https://doi.org/10.1016/j.scitotenv.2018.05.365>
- Lisboa DO, Evans HC, Araújo JPM, Elias SG, Barreto RW (2020) *Moniliophthora perniciosa*, the mushroom causing witches' broom disease of cacao: Insights into its taxonomy, ecology and host range in Brazil. *Fungal Biol* 124:983–1003. <https://doi.org/10.1016/j.funbio.2020.09.001>
- Lopes UV, Monteiro WR, Pires JL, Clement D, Yamada MM, Gramacho KP (2011) Cacao breeding in Bahia, Brazil: strategies and results. *Theor Appl Genet* 11:73–81
- Lopes UV, Pires JL, Gramacho KP, Grattapaglia D (2022) Genome-wide SNP genotyping as a simple and practical tool to accelerate the development of inbred lines in outbred tree species: An example in cacao (*Theobroma cacao* L.). *PLoS ONE* 17:e0270437. <https://doi.org/10.1371/journal.pone.0270437>
- López ME, Ramírez OA, Dubón A, Ribeiro THC, Díaz FJ, Chalfun-Junior A (2021) Sexual compatibility in cacao clones drives arrangements in the field leading to high yield. *Scientia Horticulturae* 287:110276. <https://doi.org/10.1016/j.scienta.2021.110276>
- McElroy MS, Navarro AJR, Mustiga G, Stack C, Gezan S, Peña G, Sarabia W, Saquicela D, Sotomayor I, Douglas GM, Migicovsky Z, Amores F, Tarqui O, Myles S, Motamayor JC (2018) Prediction of cacao (*Theobroma cacao*) resistance to *Moniliophthora* spp. diseases via genome-wide association analysis and genomic selection. *Front Plant Sci* 9:343. <https://doi.org/10.3389/fpls.2018.00343>
- Meinhardt LW, Rincones J, Bailey BA, Aime MC, Griffith GW, Zhang D, Pereira GAG (2008) *Moniliophthora perniciosa*, the causal agent of witches' broom disease of cacao: what's new from this old foe? *Mol Plant Pathol* 9:577–588. <https://doi.org/10.1111/j.1364-3703.2008.00496.x>

- Morrissey J, Stack JC, Valls R, Motamayor JC (2019) Low-cost assembly of a cacao crop genome is able to resolve complex heterozygous bubbles. *Hortic Res* 6:44. <https://doi.org/10.1038/s41438-019-0125-7>
- Motamayor JC, Lachenaud P, da Silva E Mota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PLoS ONE* 3:e3311
- Motamayor JC, Mockaitis K, Schmutz J, Haiminen N, Livingstone D, Cornejo O, Findley SD, Zheng P, Utro F, Royaert S, Saski C, Jenkins J, Podicheti R, Zhao M, Scheffler BE, Stack JC, Feltus FA, Mustiga GM, Amores F, Phillips W, Marelli JP, May GD, Shapiro H, Ma J, Bustamante CD, Schnell RJ, Main D, Gilbert D, Parida L, Kuhn DN (2013) The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biol* 14:r53. <https://doi.org/10.1186/gb-2013-14-6-r53>
- Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* (Edinb) 89:380–386. <https://doi.org/10.1038/sj.hdy.6800156>
- Motilal LA, Zhang D, Mischke S, Meinhardt LW, Boccara M, Fouet O, Lanaud C, Umaharan P (2016) Association mapping of seed and disease resistance traits in *Theobroma cacao* L. *Planta* 244:1265–1276. <https://doi.org/10.1007/s00425-016-2582-7>
- Mournet P, Albuquerque PSB de, Alves RM, Silva-Werneck JO, Rivallan R, Marcellino LH, Clément D (2020) A reference high-density genetic map of *Theobroma grandiflorum* (Willd. ex Spreng) and QTL detection for resistance to witches' broom disease (*Moniliophthora perniciosa*). *Tree Genetics & Genomes* 16. <https://doi.org/10.1007/s11295-020-01479-3>
- Nieves-Orduña HE, Müller M, Krutovsky KV, Gailing O (2021) Geographic patterns of genetic variation among cacao (*Theobroma cacao* L.) populations based on chloroplast markers. *Diversity* 13:249. <https://doi.org/10.3390/d13060249>
- Nieves-Orduña HE, Krutovsky KV, Gailing O (2023) Geographic distribution, conservation, and genomic resources of cacao *Theobroma cacao* L. *Crop Sci.* 63:1750–1778. <https://doi.org/10.1002/csc2.20959>
- Ofori A, Padi FK, Amoako-Attah I (2020) Field evaluation of cacao progenies derived from Guiana clones for yield and black pod disease resistance. *Crop Sci.* 60:249–261. <https://doi.org/10.1002/csc2.20101>
- Orozco-Aguilar L, López-Sampson A, Leandro-Muñoz ME, Robiglio V, Reyes M, Bordeaux M, Sepúlveda N, Somarriba E (2021) Elucidating pathways and discourses linking cocoa cultivation to deforestation, reforestation, and tree cover change in Nicaragua and Peru. *Front. Sustain. Food Syst.* 5. <https://doi.org/10.3389/fsufs.2021.635779>
- Osorio Montoya T, Henao Ramírez AM, La Hoz Vasquez T de, Urrea Trujillo AI (2022) Propagation of IMC67 plants, universal cacao (*Theobroma Cacao* L.) rootstock via somatic embryogenesis. *International Journal of Fruit Science* 22:78–94. <https://doi.org/10.1080/15538362.2021.2023067>
- Osorio-Guarín JA, Berdugo-Cely JA, Coronado-Silva RA, Baez E, Jaimes Y, Yockteng R (2020) Genome-wide association study reveals novel candidate genes associated with productivity and disease resistance to *Moniliophthora* spp. in cacao (*Theobroma cacao* L.). *G3* (Bethesda) 10:1713–1725. <https://doi.org/10.1534/g3.120.401153>
- Paulin D, Ducamp M, Lachenaud P (2008) New sources of resistance to *Phytophthora megakarya* identified in wild cocoa tree populations of French Guiana. *Crop Protection* 27:1143–1147. <https://doi.org/10.1016/j.cropro.2008.01.004>
- Peakall R, Smouse PE (2012) GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>

- Pereira AS, Lopes UV, Pires JL, Barreto MA, Pereira dos Santos L, Gonzaga RS, Da Souza Silva L, Jesus de Andrade F, Solís Bonilla JL, Gramacho KP (2021) Quantitative resistance to witches' broom disease in progenies of different sources of resistance. *Crop Protection* 146:105678. <https://doi.org/10.1016/j.cropro.2021.105678>
- Phillips-Mora W, Castillo J, Krauss U, Rodriguez E, Wilkinson MJ (2005) Evaluation of cacao (*Theobroma cacao*) clones against seven Colombian isolates of *Moniliophthora roreri* from four pathogen genetic groups. *Plant Pathology* 54:483–490. <https://doi.org/10.1111/j.1365-3059.2005.01210.x>
- Phillips-Mora W, Arciniegas-Leal A, Mata-Quirós A, Motamayor JC (2013) Catalogue of cacao clones: Selected by CATIE for commercial plantings. Technical series, technical manual, Nr. 105. CATIE, Turrialba, Costa Rica
- Ploetz R (2016) The impact of diseases on cacao production: A global overview. In: Bailey BA, Meinhardt LW (eds) *Cacao Diseases*. Springer International Publishing, Cham, pp 33–59
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Romero-Navarro JA, Phillips-Mora W, Arciniegas-Leal A, Mata-Quirós A, Haiminen N, Mustiga G, Livingstone D, van Bakel H, Kuhn DN, Parida L, Kasarskis A, Motamayor JC (2017) Application of genome wide association and genomic prediction for improvement of cacao productivity and resistance to black and frosty pod diseases. *Front Plant Sci* 8:1905
- Royaert S, Jansen J, Silva DV, Jesus Branco SM de, Livingstone DS, Mustiga G, Marelli J-P, Araújo IS, Correa RX, Motamayor JC (2016) Identification of candidate genes involved in Witches' broom disease resistance in a segregating mapping population of *Theobroma cacao* L. in Brazil. *BMC Genomics* 17:107
- Schnell RJ, Kuhn DN, Brown JS, Olano CT, Phillips-Mora W, Amores FM, Motamayor JC (2007) Development of a marker assisted selection program for cacao. *Phytopathology* 97:1664–1669. <https://doi.org/10.1094/PHTO-97-12-1664>
- Sereno ML, Albuquerque PSB, Vencovsky R, Figueira A (2006) Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. *Conserv Genet* 7:13–24. <https://doi.org/10.1007/s10592-005-7568-0>
- Singh J, Das S, Jagadis Gupta K, Ranjan A, Foyer CH, Thakur JK (2023) Physiological implications of SWEETs in plants and their potential applications in improving source-sink relationships for enhanced yield. *Plant Biotechnol J* 21:1528–1541. <https://doi.org/10.1111/pbi.13982>
- Solís Bonilla JL, Vanderlei Lopes U, Zamarripa Colmenero A, Martínez Valencia BB, Avedaño Arrazate CH, Chia Wong JA, Peres Gramacho K (2022) Path analyses define criteria that allow to reduce costs in a breeding population of cacao (*Theobroma cacao* L.). *Tree Genetics & Genomes* 18. <https://doi.org/10.1007/s11295-022-01554-x>
- Sosso D, Luo D, Li Q-B, Sasse J, Yang J, Gendrot G, Suzuki M, Koch KE, McCarty DR, Chourey PS, Rogowsky PM, Ross-Ibarra J, Yang B, Frommer WB (2015) Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. *Nat Genet* 47:1489–1493. <https://doi.org/10.1038/ng.3422>
- Thomas E, van Zonneveld M, Loo J, Hodgkin T, Galluzzi G, van Etten J (2012) Present spatial diversity patterns of *Theobroma cacao* L. in the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. *PLoS ONE* 7:e47676. <https://doi.org/10.1371/journal.pone.0047676>

- Toxopeus H (1969) Cacao (*Theobroma cacao* L.). In: Ferwerda F, Wit F (eds) Outlines of perennial crop breeding in the tropics. Veenman, Wageningen
- Turnbull C, Hadley P (2023) International Cocoa Germplasm Database (ICGD). [Online Database]. <https://www.icgd.reading.ac.uk/index.php>. Accessed 24 November 2019
- Xie H, Wang D, Qin Y, Ma A, Fu J, Qin Y, Hu G, Zhao J (2019) Genome-wide identification and expression analysis of SWEET gene family in *Litchi chinensis* reveal the involvement of LcSWEET2a/3b in early seed development. *BMC Plant Biol* 19:499. <https://doi.org/10.1186/s12870-019-2120-4>
- Zarrillo S, Gaikwad N, Lanaud C, Powis T, Viot C, Lesur I, Fouet O, Argout X, Guichoux E, Salin F, Solorzano RL, Bouchez O, Vignes H, Severts P, Hurtado J, Yopez A, Grivetti L, Blake M, Valdez F (2018) The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nat Ecol Evol* 2:1879–1888. <https://doi.org/10.1038/s41559-018-0697-x>
- Zhang D, Motilal L (2016) Origin, dispersal, and current global distribution of cacao genetic diversity. In: Bailey BA, Meinhardt LW (eds) Cacao Diseases. Springer International Publishing, Cham, pp 3–31
- Zhang D, Figueira A, Motilal L, Lachenaud P, Meinhardt LW (2011) *Theobroma*. In: Kole C (ed) Wild Crop Relatives: Genomic and Breeding Resources, vol 6. Springer, Berlin, Heidelberg, pp 277–296
- Zhang D, Martínez WJ, Johnson ES, Somarriba E, Phillips-Mora W, Astorga C, Mischke S, Meinhardt LW (2012) Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. *Genet Resour Crop Evol* 59:239–252. <https://doi.org/10.1007/s10722-011-9680-y>
- Zhen Q, Fang T, Peng Q, Liao L, Zhao L, Owiti A, Han Y (2018) Developing gene-tagged molecular markers for evaluation of genetic association of apple SWEET genes with fruit sugar accumulation. *Hortic Res* 5:14. <https://doi.org/10.1038/s41438-018-0024-3>

GENERAL DISCUSSION

The results in Chapter 2 demonstrated that cacao has a wide distribution in Amazonia (Chapter 2, Fig. 1). The majority of observation points are concentrated in the northwestern Amazon, the hotspot of cacao genetic diversity (Motamayor et al. 2008). Within this region, an area with low observation of cacao was identified, but expected cacao diversity is high (Zarrillo et al. 2018). This finding suggests that this area holds potential for future cacao collections, as illustrated in Chapter 2, Figure 1. In addition, the potential geographic distribution of cacao in Amazonia suggests that further diversity analysis using new and distinctive wild germplasm will likely reveal new genetic clusters (Cornejo et al. 2018; Zhang et al. 2011). From the observed wide geographic distribution of cacao in Amazonia, only two areas (precipitation \leq 1000 mm/year) were identified for collecting wild cacao with potential drought tolerance traits (Chapter 2, Fig. 2). These populations should be included in breeding programs in important cacao countries such as Ghana and Ivory Coast where long dry periods are affecting yield (Dzandu et al. 2021; Schroth et al. 2016)

Chapter 2 also shows that fronts of deforestation in Amazonia (Pacheco et al. 2021) are eroding cacao natural habitats, limiting the availability of diverse genetic resources for *ex situ* conservation efforts (CacaoNet 2012). In addition, cacao expeditions for *ex situ* conservation focused on searching and collecting trees with disease resistance to WBD (Zhang and Motilal 2016). In this context, Chapter 2 highlights the lack of cacao expeditions targeting wild genotypes exhibiting resistance to the economically important FPR disease (Phillips-Mora et al. 2005).

The review of the genetic basis of agronomic traits such as disease resistance, yield and quality traits (flavor and fat) suggests a polygenic nature (Chapter 2). For example, there are at least 44 genes associated with disease resistance for prospective genetic improvement of cacao (Chapter 2, Table 1). Modern tree breeding programs, instead of relying on major QTL, have

adopted genomic selection, that can predict complex traits using thousands of genome-wide SNP markers, accelerating breeding cycles and increasing selection intensity (Grattapaglia et al. 2018; Grattapaglia 2022). In cacao, genomic selection has demonstrated to improve selection since individual effects of multiple SNP markers are considered in complex cacao traits such as yield and disease resistance (Bekele et al. 2022; McElroy et al. 2018; Romero-Navarro et al. 2017).

The high chloroplast diversity observed in northwestern Amazonia highlights this area as a center of cacao genetic diversity (Chapter 3). The observation of 23 chloroplast haplotypes including seven geographically restricted haplotypes suggests that *in situ* conservation areas can be proposed in northwestern Amazon (Chapter 3, Fig 2-3). Previous studies of tropical tree species such as Cedar (*Cedrella odorata*) used the geographic distribution of chloroplast haplotypes to propose areas for *in situ* conservation in Central America (Cavers et al. 2003). The pattern of dispersal of chloroplast haplotypes observed from the west to east Amazon suggests a decline in cacao genetic diversity (Chapter 3, Fig 2-3). This geographic pattern of decline in cacao genetic diversity was also observed by Cornejo et al. (2018). This could be explained by past human influences, as cacao was spread in Amazonia by humans, as it was the case with other tropical tree species with crop potential (Clement 1999; Levis et al. 2017).

The structure analysis of cacao germplasm using SNPs associated with agronomic traits identified four genetic clusters reflecting the geography origin of samples (Chapter 4, Figure 1). Managed cacao (breeding populations and cultivated varieties) was represented mainly by two genetic clusters, suggesting a low population structure. This germplasm structure supports the findings of Cornejo et al. (2018). According to their research, modern man-made cacao hybrids are largely represented by two out of ten identified cacao genetic clusters (Cornejo et al. 2018; Motamayor et al. 2008). These results suggest the opportunity to broaden the genetic base of cultivated cacao which is hampered by diseases and low yield (Bekele and Phillips-

Mora 2019; Gutiérrez et al. 2016). For example, Guiana germplasm was observed in low proportion in managed cacao but previous agronomic evaluations suggest that disease resistance and yield traits can be exploited by cacao breeding programs (Lachenaud et al. 2007). Divergence of wild and managed cacao ($F_{ST} > 0.05$) suggest selection processes in cacao germplasm in agronomic traits such disease resistance (WBD and CW) and yield (pod index and high yield) (Chapter 4, Table 3). These patterns are associated with historical process in cacao selection and cultivation (Toxopeus 1969; Zhang et al. 2011)

The heterozygous genotypes observed in cacao accessions with disease resistance traits suggest QTL that potentially contribute to cacao heterosis. For example, the genotypes CT (SNP Tcm004s00110232) related to witches' broom disease (Royaert et al. 2016) and the genotype GT (SNP Tcm006s13222057) related to Ceratosystis disease (Fernandes et al. 2018) were observed in accessions reported as disease resistant-tolerant (Chapter 4, Fig 5-6). Validation of SNP markers associated with agronomic traits should be validated for different cacao breeding programs to facilitate the identification QTL for cultivar development.

REFERENCES

- Bekele, F., Bidaisee, G., Allegre, M., Argout, X., Fouet, O., Boccara, M., Saravanakumar, D., Bekele, I., & Lanaud, C. (2022). Genome-wide association studies and genomic selection assays made in a large sample of cacao (*Theobroma cacao* L.) germplasm reveal significant marker-trait associations and good predictive value for improving yield potential. *PLoS One*, *17*(10), e0260907. <https://doi.org/10.1371/journal.pone.0260907>
- Bekele, F., & Phillips-Mora, W. (2019). Cacao (*Theobroma cacao* L.) breeding. In J. M. Al-Khayri, S. M. Jain, & D. V. Johnson (Eds.), *Advances in Plant Breeding Strategies: Industrial and Food Crops* (Vol. 46, pp. 409–487). Springer International Publishing. https://doi.org/10.1007/978-3-030-23265-8_12
- CacaoNet. (2012). *A global strategy for the conservation and use of cacao genetic resources, as the foundation for a sustainable cocoa economy*. Montpellier, France. Bioversity International.
- Cavers, S., Navarro, C., & Lowe, A. J. (2003). Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, *Cedrela odorata* L., in Mesoamerica. *Molecular Ecology*, *12*(6), 1451–1460. <https://doi.org/10.1046/j.1365-294X.2003.01810.x>
- Clement, C. R. (1999). 1492 and the loss of Amazonian crop genetic resources. ii. Crop biogeography at contact. *Economic Botany*, *53*(2), 203–216. <https://doi.org/10.1007/BF02866499>
- Cornejo, O. E., Yee, M.-C., Dominguez, V., Andrews, M., Sockell, A., Strandberg, E., Livingstone, D., Stack, C., Romero, A., Umaharan, P., Royaert, S., Tawari, N. R., Ng, P., Gutierrez, O., Phillips, W., Mockaitis, K., Bustamante, C. D., & Motamayor, J. C. (2018). Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process. *Communications Biology*, *1*, 167. <https://doi.org/10.1038/s42003-018-0168-6>
- Dzandu, E., Enu-Kwesi, L., Markwei, C. M., & Ayeh, K. O. (2021). Screening for drought tolerance potential of nine cocoa (*Theobroma cacao* L.) genotypes from Ghana. *Heliyon*, *7*(11), e08389. <https://doi.org/10.1016/j.heliyon.2021.e08389>
- Fernandes, L. D. S., Royaert, S., Corrêa, F. M., Mustiga, G. M., Marelli, J.-P., Corrêa, R. X., & Motamayor, J. C. (2018). Mapping of a major QTL for *Ceratocystis* Wilt disease in an F1 population of *Theobroma cacao*. *Frontiers in Plant Science*, *9*, 155. <https://doi.org/10.3389/fpls.2018.00155>
- Grattapaglia, D. (2022). Twelve years into genomic selection in forest trees: climbing the slope of enlightenment of marker assisted tree breeding. *Forests*, *13*(10), 1554. <https://doi.org/10.3390/f13101554>
- Grattapaglia, D., Silva-Junior, O. B., Resende, R. T., Cappa, E. P., Müller, B. S. F., Tan, B., Isik, F., Ratcliffe, B., & El-Kassaby, Y. A. (2018). Quantitative Genetics and Genomics Converge to Accelerate Forest Tree Breeding. *Frontiers in Plant Science*, *9*, 1693. <https://doi.org/10.3389/fpls.2018.01693>
- Gutiérrez, O., Campbell, A., & Phillips-Mora, W. (2016). Breeding for disease resistance in cacao. In B. A. Bailey & L. W. Meinhardt (Eds.), *Cacao Diseases* (pp. 567–609). Springer International Publishing. https://doi.org/10.1007/978-3-319-24789-2_18

- Lachenaud, P., Paulin, D., Ducamp, M., & Thevenin, J.-M. (2007). Twenty years of agronomic evaluation of wild cocoa trees (*Theobroma cacao* L.) from French Guiana. *Scientia Horticulturae*, 113(4), 313–321. <https://doi.org/10.1016/j.scienta.2007.05.016>
- Levis, C., Costa, F. R. C., Bongers, F., Peña-Claros, M., Clement, C. R., Junqueira, A. B., Neves, E. G., Tamanaha, E. K., Figueiredo, F. O. G., Salomão, R. P., Castilho, C. V., Magnusson, W. E., Phillips, O. L., Guevara, J. E., Sabatier, D., Molino, J.-F., López, D. C., Mendoza, A. M., Pitman, N. C. A., . . . Steege, H. (2017). Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science (New York, N.Y.)*, 355(6328), 925–931. <https://doi.org/10.1126/science.aal0157>
- McElroy, M. S., Navarro, A. J. R., Mustiga, G., Stack, C., Gezan, S., Peña, G., Sarabia, W., Saquicela, D., Sotomayor, I., Douglas, G. M., Migicovsky, Z., Amores, F., Tarqui, O., Myles, S., & Motamayor, J. C. (2018). Prediction of cacao (*Theobroma cacao*) resistance to *Moniliophthora* spp. Diseases via genome-wide association analysis and genomic selection. *Frontiers in Plant Science*, 9, 343. <https://doi.org/10.3389/fpls.2018.00343>
- Motamayor, J. C., Lachenaud, P., da Silva E Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PloS One*, 3(10), e3311. <https://doi.org/10.1371/journal.pone.0003311>
- Pacheco, P., Mo, K., Dudley, N., Shapiro, A., Aguilar-Amuchastegui, N., Ling, P., Anderson, C., & Marx, A. (2021). *Deforestation fronts: Drivers and responses in a changing world*. Gland, Switzerland. WWF.
- Phillips-Mora, W., Castillo, J., Krauss, U., Rodriguez, E., & Wilkinson, M. J. (2005). Evaluation of cacao (*Theobroma cacao*) clones against seven Colombian isolates of *Moniliophthora roreri* from four pathogen genetic groups. *Plant Pathology*, 54(4), 483–490. <https://doi.org/10.1111/j.1365-3059.2005.01210.x>
- Romero-Navarro, J. A., Phillips-Mora, W., Arciniegas-Leal, A., Mata-Quirós, A., Haiminen, N., Mustiga, G., Livingstone, D., van Bakel, H., Kuhn, D. N., Parida, L., Kasarskis, A., & Motamayor, J. C (2017). Application of genome wide association and genomic prediction for improvement of cacao productivity and resistance to black and frosty pod diseases. *Frontiers in Plant Science*, 8, 1905. <https://doi.org/10.3389/fpls.2017.01905>
- Royaert, S., Jansen, J., Silva, D. V., Jesus Branco, S. M. de, Livingstone, D. S., Mustiga, G., Marelli, J.-P., Araújo, I. S., Correa, R. X., & Motamayor, J. C (2016). Identification of candidate genes involved in Witches' broom disease resistance in a segregating mapping population of *Theobroma cacao* L. In Brazil. *BMC Genomics*, 17, 107.
- Schroth, G., Läderach, P., Martinez-Valle, A. I., Bunn, C., & Jassogne, L. (2016). Vulnerability to climate change of cocoa in West Africa: Patterns, opportunities and limits to adaptation. *The Science of the Total Environment*, 556, 231–241. <https://doi.org/10.1016/j.scitotenv.2016.03.024>
- Toxopeus, H. (1969). Cacao (*Theobroma cacao* L.). In F. Ferwerda & F. Wit (Eds.), *Outlines of perennial crop breeding in the tropics*. Veenman.
- Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I., Fouet, O., Argout, X., Guichoux, E., Salin, F., Solorzano, R. L., Bouchez, O., Vignes, H., Severt, P., Hurtado, J., Yepez, A., Grivetti, L., Blake, M., & Valdez, F. (2018). The use and

domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nature Ecology & Evolution*, 2(12), 1879–1888. <https://doi.org/10.1038/s41559-018-0697-x>

Zhang, D., Figueira, A., Motilal, L., Lachenaud, P [Philippe], & Meinhardt, L. W. (2011). *Theobroma*. In C. Kole (Ed.), *Wild Crop Relatives: Genomic and Breeding Resources* (Vol. 6, pp. 277–296). Springer. https://doi.org/10.1007/978-3-642-21201-7_13

Zhang, D., & Motilal, L. (2016). Origin, dispersal, and current global distribution of cacao genetic diversity. In B. A. Bailey & L. W. Meinhardt (Eds.), *Cacao Diseases* (pp. 3–31). Springer International Publishing. https://doi.org/10.1007/978-3-319-24789-2_1

CONCLUSIONS

The conservation and use of cacao genetic diversity in breeding programs are vital for sustainable cacao cultivation and the global chocolate industry. The geographic distribution of cacao pinpoints to new areas for germplasm collection and supports northwestern Amazon as the hotspot of cacao genetic diversity. This was also validated by the observation of areas with high haplotype diversity in the Peruvian Amazon. In addition, observations of wild cacao in Amazonia indicate the existence of natural populations that have not yet been subjected to population genomic analysis, and have not been included in breeding efforts. New genetic clusters with allelic variants that influence agronomic traits can be identified. However, trends in Amazon deforestation limit the conservation and use of cacao primary gene pool.

Cacao germplasm (cultivated and breeding populations) shows a narrow genetic diversity reflecting historical processes of selection. Agronomic traits that distinguish wild and managed cacao include disease resistance to WBD and yield traits. Germplasm resources observed with low contribution to managed cacao but with agronomic potential include the Guiana germplasm. Cacao SNPs linked to agronomic traits need to be validated under different environments to facilitate the identification of major QTL for cultivar development.

Overall, cultivated cacao can benefit from novel genetic resources available in wild germplasm or obtained from new collections in northwestern Amazonia. Standardized agronomic and genotypic evaluation add value to cacao genetic resources held in *ex situ* collections and improve its utilization in breeding programs. Challenges remain for the pyramiding of agronomic traits in cacao, such as developing inbred lines to exploit heterosis and genomic selection in important agronomic traits such as disease resistance to BP, WBD and FPR, yield and seed quality traits.

OUTLOOK

Based on the results presented in chapters 2, 3, and 4, future research areas within the frame of cacao genetic diversity and agronomic traits that will contribute to building breeding populations and ultimately developing cultivars are presented.

Population differentiation studies using neutral molecular markers are likely to reveal new cacao clusters, and forest areas in northwestern Amazonia with potential for collecting diverse germplasm are presented in Fig. 1, Chapter 2. In addition, collecting and conserving diverse cacao germplasm from wild habitats will also increase the availability of genetic resources for breeding and help to understand how cacao genetic diversity varies geographically and is distributed across Amazonia. In this context, further analysis with more diverse samples and informative cacao chloroplast SSR markers (Chapter 3) will help to better understand the phylogeographic structure of cacao in Amazonia. Unfortunately, the natural habitat of cacao is disappearing due to deforestation (Fig. 1, Chapter 2), hindering future research progress in population genetic analysis and cacao germplasm conservation.

Cacao germplasm conserved in *ex situ* collections should be analyzed with SNP markers to identify genetic gaps, mislabeling and redundancy, followed by standard agronomic evaluations to establish an efficient conservation and increased use of cacao genetic diversity in cultivar development (Chapter 2). These analyses are critical to increase the genetic diversity of cultivated cacao, as we observed two overrepresented genetic clusters among cultivars and breeding populations (Chapter 4). Novel genetic resources are needed to broaden the genetic base of breeding populations in support of agronomic traits. For example, the agronomic potential of Guiana germplasm observed in less proportion in breeding and cultivated cacao should be evaluated under different environments (Chapter 4).

Research on traits of economic importance includes identification of novel alleles associated with resistance to different pathogen strains of WBD and FPR (Chapter 2). These

studies will benefit by using diverse cacao germplasm (Chapter 4) and standard phenotyping allowing comparability of genotype performance between breeding programs.

Introgressions of quality traits in cacao such as low seed Cd accumulation are needed in cacao cultivars (Chapter 2). Identification of genotypes with low seed Cd accumulation are needed to select potential rootstocks. The breeding challenge is to combine low Cd with disease resistance traits in cacao germplasm used as rootstock (e.g., IMC 67). Cacao candidate genes associated with Cd transportation for perspective genetic improvement of cacao include *TcNRAMP5*, *TcHMA3*, and *TcHMA2*. Thanks to advances in cacao gene editing, these genes can be targeted to reduce seed Cd accumulation. Still, it requires high-quality reference genomes and protocols for somatic embryogenesis in cacao cultivars (Chapter 2).

Drought tolerance is a trait not well studied in cacao breeding programs. The identified natural populations growing in areas with precipitations ≤ 1000 mm/year (Fig. 2, Chapter 2) can help to build breeding populations to increase drought tolerance by targeting root traits such as angle and root biomass that improve water use efficiency.

Although there are advances in developing major QTL for cacao breeding (Chapter 4), future research areas in cacao breeding should implement genomic selection. For example, development of cacao inbred lines to exploit heterosis followed by genomic selection of untested clones under different environments should be performed. Genomic resources such as the cacao Infinium array (15K SNPs) should be used by breeding programs in analytical and experimental studies such as population differentiation, association studies and genomic selection.

SUMMARY

The tropical cacao tree, *Theobroma cacao* L., is cultivated to produce seeds, the unique raw material for the chocolate industry. Thus, conservation and use of cacao genetic resources in breeding programs are vital for the cacao-chocolate global economy. Wild cacao populations have a wide geographic distribution in their native range along the Amazon basin, but northwestern Amazonia is the hotspot of cacao genetic diversity. From the rich diversity of cacao (at least twelve genetic clusters are identified) the domestication of cacao in South America led to the development of Criollo, whose seeds are characterized by reduced bitterness, a trait appreciated for the international cacao market. Criollo ancestry still is a primary source for the development of cultivated cacao varieties. A narrow genetic base in cultivated cacao and disease vulnerability led to collecting disease resistant genotypes in Amazonia since the 1930s. This was followed by conservation via *ex situ* collections, agronomic evaluation of accessions to improve and broaden the genetic base of cultivated cacao. Disease resistance and yield are the most important economic traits in cocoa breeding, as diseases can destroy up to 40% of cocoa production. The conservation and utilization of the rich cacao diversity are the primary source of novel alleles necessary to address current and future cacao breeding objectives.

The first paper reviews key topics in cacao genetics essential for crop improvement. By analyzing the geographic distribution of wild cacao in Amazonia, which consisted of 4413 observation points compiled from the years 1938 to 2022, and mapping the patterns of genetic variation, new areas for germplasm collection were identified in the border region between Ecuador and Peru in northwestern Amazonia. Despite the deforestation in Amazonia, this analysis suggests that it is still possible to collect wild material in the identified areas. In terms of conservation, germplasm collections in cacao producing countries can benefit from available

SNP markers to improve genotype identity and characterization of agronomic traits at early stages.

The studies addressing the genetic basis of disease resistance, seed quality traits, and yield suggest a polygenic architecture. For example, at least 44 genes reported and distributed over the ten cacao chromosomes are associated with resistance to three different diseases (Black Pod-BP, Witches' Broom Disease WBD and Frosty Pod Rot-FPR). Further evaluations with more diverse germplasm and different fungal strains likely will reveal more disease resistance genes and genotypes. Recently, the genetic basis of seed quality traits such as reduced cadmium and flavor has been addressed, but informative molecular markers need to be developed to accelerate breeding. Other traits of particular importance to West Africa, the largest producer of cacao, include heat and drought tolerance. Cacao breeding efforts have not fully exploited heterosis but can profit from modern genome-wide SNP data to develop inbred lines and design crosses, followed by genomic selection of untested hybrids. Overall, cacao breeding requires data generated by standard phenotyping (of different cacao populations) integrated with analytical methods such as genome-wide association studies (GWAS), quantitative trait locus (QTL) mapping, genomic selection, and field trials to identify relevant genotypes for breeding and cultivation.

The second paper investigates the phylogeographic structure of cacao using nine informative chloroplast DNA (cpDNA) microsatellite (simple sequence repeat - SSR) markers to genotype 233 cacao samples. Nineteen out of the twenty-three chloroplast haplotypes identified were observed in western Amazonia, including seven geographically restricted haplotypes, confirming northwestern Amazonia as a hotspot for cacao genetic diversity and center of crop domestication. Two observed haplotypes were widespread from west to east along the Amazon basin, suggesting seed dispersal by humans. When comparing the observed haplotypes with previously identified genetic clusters using nuclear SSRs, specific chloroplast

haplotypes were observed within eight genetic clusters. Evidence of selection of plant material was provided by the observation of a single haplotype common in cacao farms of west Ecuador and reference Trinitario accessions, the most important breeding population. The identified chloroplast haplotypes can be used to determine the geographic origin of planting material at a finer geographic scale. In addition, these chloroplast haplotypes, in combination with phenotypic assessments, can help to select geographically distinctive varieties for cacao breeding programs.

The third paper analyses the population structure of wild and managed cacao (cultivated and breeding populations), identifies agronomic traits under selection, and uncovers valuable cacao genotypes not observed in cultivated and breeding populations. Forty published SNPs associated with disease resistance and 11 SNPs with yield traits were investigated in 346 accessions using the MassARRAY® system. The structure analysis revealed four genetic clusters. Two clusters were overrepresented in managed cacao, which is characterized by admixed individuals, representing artificial hybrids. A cluster overrepresented in managed cacao is linked to the preference for maintaining the superior flavor of cultivated cacao from the Criollo ancestry. Under-utilization of genetic resources such as Guiana cacao was detected in managed cacao, and published agronomic evaluation of this germplasm group showed yield and disease resistance traits. Yield and disease resistance traits (mainly resistance to WBD) show divergence between wild and cultivated cacao ($F_{ST} > 0.05$). This divergence likely reflects selection during domestication, cultivation, and breeding. The identified SNPs showing divergence between wild and cultivated cocoa can be used to improve selection in breeding populations. Agronomic evaluation of wild populations can expand the genetic base for breeding cacao, especially for traits of economic importance such as low seed Cadmium and disease resistance against FPR.

ZUSAMMENFASSUNG

Der tropische Kakaobaum, *Theobroma cacao* L., wird für die Samenproduktion angebaut, die den einzigartigen Rohstoff für die Schokoladenindustrie darstellen. Daher sind Zuchtprogramme für Erhaltung und Nutzung der genetischen Ressourcen von Kakao für die globale Wirtschaft und den Handel von Kakao und Schokolade von entscheidender Bedeutung. Wildkakao weist eine große geografische Verbreitung auf, die sich ursprünglich entlang des Amazonasbeckens erstreckte. Hierbei liegt der „Hotspot“ der genetischen Vielfalt von Kakao im Nordwesten Amazoniens. Insgesamt wurden in der Art zwölf genetische Gruppen (Cluster) identifiziert. Aus dieser reichen Vielfalt des Kakaos führte die Domestizierung des Kakaos in Südamerika zur Entwicklung der Sorte Criollo, dessen Samen sich durch eine reduzierte Bitterkeit auszeichnen. Diese Eigenschaft wird auf dem internationalen Kakaomarkt hochgeschätzt. Die Criollo-Abstammung ist nach wie vor eine Hauptquelle für die Entwicklung kultivierter Kakaosorten. Eine schmale genetische Basis in den kultivierten Kakaosorten und deren Anfälligkeit für Krankheiten führten dazu, dass seit den 1930er Jahren krankheitsresistente Genotypen in Amazonien gesammelt wurden. Diese Akzessionen wurden in *ex-situ*-Sammlungen erhalten und eine agronomische Bewertung durchgeführt. Sie erhielten Einzug in die Zuchtprogramme, um die genetische Basis des angebauten Kakaos zu verbessern und zu erweitern. Krankheitsresistenz und Ertrag sind die wichtigsten wirtschaftlichen Merkmale in der Kakaozüchtung, da Krankheiten bis zu 40% der Kakaoproduktion zerstören können. Die Erhaltung und Nutzung der reichen Kakaovielfalt sind die Hauptquelle für neue Allele, die notwendig sind, um aktuelle und zukünftige Ziele der Kakaozüchtung zu erreichen.

Der erste Artikel befasst sich mit Schlüsselthemen der Kakaogenetik, die für die Verbesserung von Nutzpflanzen unerlässlich sind. Durch die Analyse der geografischen Verbreitung von Wildkakao in Amazonien, die aus Beobachtungspunkten aus den Jahren 1938 bis 2022 bestand, und die Kartierung der Muster der genetischen Variation wurden neue

Gebiete für die Sammlung genetischer Ressourcen in der Grenzregion zwischen Ecuador und Peru im Nordwesten Amazoniens identifiziert. Trotz der Entwaldung in Amazonien deuten die Analysen darauf hin, dass es in den identifizierten Gebieten immer noch möglich ist, Wildmaterial zu sammeln. In Bezug auf den Naturschutz können Genbanken in kakaoproduzierenden Ländern von den verfügbaren SNP-Markern profitieren, um die Genotypidentität und die Charakterisierung agronomischer Merkmale in frühen Stadien zu verbessern. Die Studien, die sich mit den genetischen Grundlagen der Krankheitsresistenz, den Merkmalen der Samenqualität und dem Ertrag befassen, deuten auf eine polygene Architektur hin. Zum Beispiel werden 40 Gene, die über die zehn Kakaochromosomen verteilt sind, mit der Resistenz gegenüber dreier Krankheiten (Kakao- oder Schwarzfäule, Hexenbesenkrankheit und *Monilia*-Fruchtfäule) in Verbindung gebracht. Weitere Auswertungen mit genetisch vielfältigerem Pflanzenmaterial und verschiedenen Pilzstämmen werden wahrscheinlich mehr krankheitsresistente Gene und Genotypen aufdecken. In jüngster Zeit wurde die genetische Grundlage von Saatgutqualitätsmerkmalen wie reduziertem Cadmiumgehalt und geschmacksbestimmenden Inhaltsstoffen untersucht. Diese Arbeiten könnten für die Entwicklung aussagekräftiger molekulare Marker für eine Beschleunigung der Züchtung genutzt werden. Weitere Eigenschaften, die für Westafrika als den größten Kakaoproduzenten, von besonderer Bedeutung sind, sind Hitze- und Trockenheitstoleranz. Die Bemühungen in der Kakaozüchtung haben die Heterosis nicht vollständig ausgenutzt, können aber von modernen genomweiten SNP-Daten profitieren, um Inzuchtlinien zu entwickeln und Kreuzungen zu entwerfen, gefolgt von genomischer Selektion ungetesteter Hybriden. Insgesamt erfordert die Kakaozüchtung Daten, die durch Standard-Phänotypisierung (verschiedener Kakaopopulationen) generiert werden. Diese werden mit analytischen Methoden wie genomweiten Assoziationsstudien (GWAS), *Quantitative Trait Locus* (QTL)-Kartierung,

genomischer Selektion und Feldversuchen zur Identifizierung relevanter Genotypen für Züchtung und Anbau kombiniert.

Die zweite Arbeit untersucht die phylogeografische Struktur von Kakao unter Verwendung von neun informativen Mikrosatellitenmarkern (*Simple Sequence Repeat* - SSR) der Chloroplasten-DNA (cpDNA) für 233 Kakaoproben. Neunzehn der dreiundzwanzig identifizierten Chloroplasten-Haplotypen wurden im westlichen Amazonasgebiet beobachtet, darunter sieben geografisch begrenzte Haplotypen, was den Nordwesten Amazoniens als „Hotspot“ für die genetische Vielfalt von Kakao und als Zentrum der Domestizierung von Nutzpflanzen bestätigt. Zwei beobachtete Haplotypen waren von West nach Ost entlang des Amazonasbeckens weit verbreitet, was auf Samenausbreitungsmuster durch den Menschen hindeutet. Beim Vergleich der beobachteten Haplotypen mit zuvor identifizierten genetischen Clustern unter Verwendung nuklearer SSRs wurden spezifische Chloroplasten-Haplotypen innerhalb von acht genetischen Clustern beobachtet. Einen Beweis für die Selektion des Pflanzenmaterials lieferte die Beobachtung eines einzigen Haplotyps, der in den Kakaofarmen West Ecuadors und der wichtigsten Zuchtpopulation der Trinitario-Akzessionen verbreitet ist. Die identifizierten Chloroplasten-Haplotypen können verwendet werden, um die geografische Herkunft von Pflanzmaterial auf einer feineren geografischen Skala zu bestimmen. Darüber hinaus können diese Chloroplasten-Haplotypen in Kombination mit phänotypischen Bewertungen dazu beitragen, geografisch unterschiedliche Sorten für Kakaozüchtungsprogramme auszuwählen.

Die dritte Arbeit analysiert die Populationsstruktur von Wild-, Kultur- und Zuchtpopulationen. Es konnten, agronomische Merkmale unter Selektion identifiziert und wertvolle Kakao-Genotypen aufgedeckt werden, die in Kultur- und Zuchtpopulationen nicht beobachtet wurden. Vierzig publizierte SNPs assoziiert mit Krankheitsresistenz und 11 SNPs mit Ertragsmerkmalen wurden in 367 Akzessionen mit dem MassARRAY-System®

untersucht. Die Strukturanalyse ergab vier genetische Cluster. Zwei Cluster waren im bewirtschafteten Kakao überrepräsentiert, welcher durch genetisch gemischte Individuen gekennzeichnet ist, die künstliche Hybriden darstellen. Ein Cluster, der bei bewirtschaftetem Kakao überrepräsentiert ist, hängt mit der Vorliebe für die Beibehaltung des überlegenen Geschmacks von kultiviertem Kakao aus der Criollo-Abstammung zusammen. Eine Krankheitsresistenz gegen die Hexenbesenkrankheit (engl.: Witches' Broom Disease—WBD) wurde auch bei der Introgression von Wildmaterial beobachtet, das in Peru gesammelt wurde. Bei bewirtschaftetem Kakao konnte eine geringe Verwendung einiger genetischer Akzessionen wie Guyana-Kakao festgestellt werden. Diese Akzessionen zeigen in veröffentlichten agronomischen Bewertungen bessere Ertrags- und Krankheitsresistenzmerkmale gegenüber anderen Sorten/Gruppen. Ertrags- und Krankheitsresistenzmerkmale (hauptsächlich Resistenz gegen WBD) zeigen eine Divergenz zwischen Wild- und Kulturkakao ($F_{ST} > 0,05$). Diese Divergenz spiegelt wahrscheinlich die Selektion während der Domestizierung, Kultivierung und Zucht wider. Die identifizierten SNPs, die eine Divergenz zwischen Wild- und Kulturkakao zeigen, können zur Verbesserung der Selektion in Zuchtpopulationen verwendet werden. Agronomische Auswertungen von Wildpopulationen können die genetische Basis für die Zucht von Kakao erweitern, insbesondere für Merkmale von wirtschaftlicher Bedeutung wie geringer Cadmium-Konzentration in Samen und Krankheitsresistenz gegen Frosty Pod Rot (FPR).