RESEARCH ARTICLE

Climate change impact on cultivated and wild cacao in Peru and the search of climate change-tolerant genotypes

Viviana Ceccarelli[1] | Tobias Fremout[1,2] | Diego Zavaleta[1] | Sphyros Lastra[1] |
Sixto Imán Correa[3] | Enrique Arévalo-Gardini[4,5] | Carlos Armando Rodriguez[6] |
Wilbert Cruz Hilacondo[3] | Evert Thomas[1]

[1]Bioversity International, Lima, Peru

[2]Division of Forest, Nature and Landscape, KU Leuven, Leuven, Belgium

[3]Instituto Nacional de Innovación Agraria, La Molina, Peru

[4]Instituto de Cultivos Tropicales (ICT), Tarapoto, Peru

[5]Universidad Nacional Autonoma de Alto Amazonas, Yurimaguas, Peru

[6]Servicio Nacional de Sanidad y Calidad Agroalimentaria, Quillabamba, Peru

Correspondence
Evert Thomas, Bioversity International, Lima, Peru.
Email: e.thomas@cgiar.org

Abstract

Aim: Cacao (*Theobroma cacao* L.) is expected to be vulnerable to climate change. The objectives of this study were to (a) assess the future impact of climate change on cacao in Peru and (b) identify areas where climate change-tolerant genotypes are potentially present.

Location: Peru

Methods: Drawing on 19,700 and 1,200 presence points of cultivated and wild cacao, respectively, we modelled their suitability distributions using multiple ensemble models constructed based on both random and target group selection of pseudo-absence points and different resolutions of spatial filtering. To estimate the uncertainty of future predictions, we generated future projections for all the ensemble models. We investigated the potential emergence of novel climates, determined expected changes in ecogeographical zones (zones representative for particular sets of growth conditions) and carried out an outlier analysis based on the environmental variables most relevant for climate change adaptation to identify areas where climate change-tolerant genotypes are potentially present.

Results: We found that the best modelling approaches differed between cultivated and wild cacao and that the resolution of spatial filtering had a strong impact on future suitability predictions, calling for careful evaluation of the effect of model selection on modelling results. Overall, our models foresee a contraction of suitable area for cultivated cacao while predicting a more positive future for wild cacao in Peru. Ecogeographical zones are expected to change in 8%–16% of the distribution of cultivated and wild cacao. We identified several areas where climate change-tolerant genotypes may be present in Peru.

Main conclusions: Our results indicate that tolerant genotypes will be required to facilitate the adaptation of cacao cultivation under climate change. The identified cacao populations will be target of collection missions.

1 | INTRODUCTION

Climate change is expected to severely impact agricultural production. Global average temperature has already increased by about 0.85°C during the period 1880–2012 and is projected to further increase by 1.4–3.1°C towards the end of this century (Stocker et al., 2013). Rising temperatures in combination with changes in precipitation patterns and higher frequencies of extreme droughts and floods (Cai et al., 2014; Stocker et al., 2013) are expected to reduce crop yields and threaten farmers' livelihoods around the world (Parry et al., 2004; Porter et al., 2014). Impacts are likely to disproportionately affect agricultural systems in the tropics, both because negative effects on yields are expected to be larger compared with temperate zones (Parry et al., 2004; Porter et al., 2014) and because smallholder systems have lower resilience due to their limited available resources (Morton, 2007). As climate change will have differential impacts on individual crop species or varieties and growing areas (Porter et al., 2014), urgent action is needed to better understand the expected impacts of climate change on specific crops, particularly in the tropics, to guide the development of adaptation strategies.

Cacao (*Theobroma cacao* L.), one of the most important cash crops in the tropics, is expected to be vulnerable to climate change (Läderach et al., 2013; Lahive et al., 2019; Medina & Laliberte, 2017). Evolved as an understory tree native to the Amazon basin, cacao is a drought-sensitive species and high temperatures and low precipitation can negatively impact its growth, productivity and yield quality (Daymond & Hadley, 2008; Moser et al., 2010; Zuidema et al., 2005). Increasing evidence indicates that climate change is already affecting cacao cultivation around the world, resulting in higher mortality, declines in yield quality and quantity (Medina & Laliberte, 2017), and increased incidence of diseases (Gateau-Rey et al., 2018).

Breeding for climate change tolerance represents one of the most promising adaptation options to climate change (Dempewolf et al., 2014; Smit & Skinner, 2002). The wide genetic diversity of cacao represents a key resource to develop climate change-tolerant varieties (Lahive et al., 2019; Medina & Laliberte, 2017). For instance, numerous studies have reported high variability among cacao genotypes in response to elevated temperatures (Daymond & Hadley, 2004, 2008), water deficit (Apshara et al., 2016; Araque et al., 2012; Ofori et al., 2014; dos Santos et al., 2014) and flooding (Bertolde et al., 2009, 2012; De Almeida et al., 2016). However, despite its great potential, cacao's full genetic diversity has thus far been largely unexploited by breeders (Lahive et al., 2019; Medina & Laliberte, 2017) and climate change-tolerant cacao varieties are still lacking to date (Farrell et al., 2018; Lahive et al., 2019; Medina & Laliberte, 2017).

Habitat suitability models—also known as species distribution or ecological niche models—are useful tools to both assess the future impact of climate change and to guide the selection of climate change-tolerant genotypes. These models can estimate current and future suitability distributions of species and describe their environmental range and can therefore be used to identify regions where genotypes adapted to climatic extremes are potentially present. In the case of cacao, recent suitability modelling studies have predicted a reduction in the growing area and a shift towards higher elevations. However, these studies have focussed on West Africa (Läderach et al., 2013; Schroth et al., 2016) and Central America (Eitzinger et al., 2015; de Sousa et al., 2019), and until today research in the centre of origin of cacao remains limited. To date, suitability modelling has not been used to support genotype selection yet (Medina & Laliberte, 2017).

The objectives of this paper were to (a) assess the future impact of climate change on the distribution of suitable habitat of cultivated and wild cacao in Peru and (b) identify areas where climate change-tolerant genotypes are potentially present. Peru has great potential for cacao genotype selection as it presents high levels of genetic diversity and is the country of origin of multiple genetic groups recognized within the species (Motamayor et al., 2008; Thomas et al., 2012). Because exchange of materials between countries is still restricted, and in order to identify locally adapted genotypes in the country, the search of genotypes is limited within Peru. We used an ensemble modelling approach to assess the distribution of suitable habitat under current and future climatic conditions considering three time horizons (2030s, 2050s, 2070s) and two greenhouse gas emission scenarios (RCP4.5 and RCP8.5). Because model selection can significantly influence model results (Araújo & New, 2007; Thuiller et al., 2019), we assessed the impact of different approaches of model assembly on current and future suitability distributions. Furthermore, we investigated the potential emergence of novel climates within the cacao suitable area and determined how cacao ecogeographical zones (representative for particular sets of growth conditions) might change under future climates. Finally, we carried out an outlier analysis based on the environmental variables most relevant for climate change adaptation to identify areas where climate change-tolerant genotypes are potentially present, which will be target of collection missions.
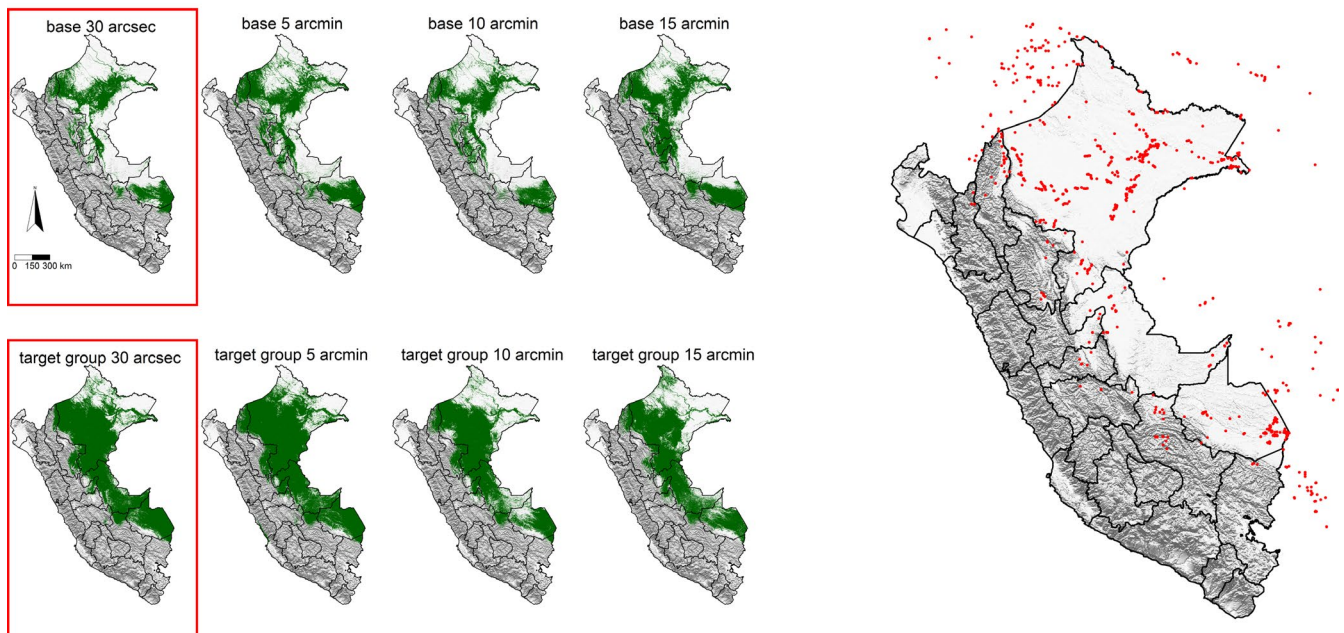
2 | METHODS

2.1 | Presence data

Cacao presence points were compiled from several sources (Text S1 in Appendix S1) and divided into cultivated and wild cacao according to the indication in the source. For cultivated cacao, we retained only presence points within Peru (Figure 1a). For wild cacao,

(a) Cultivated cacao
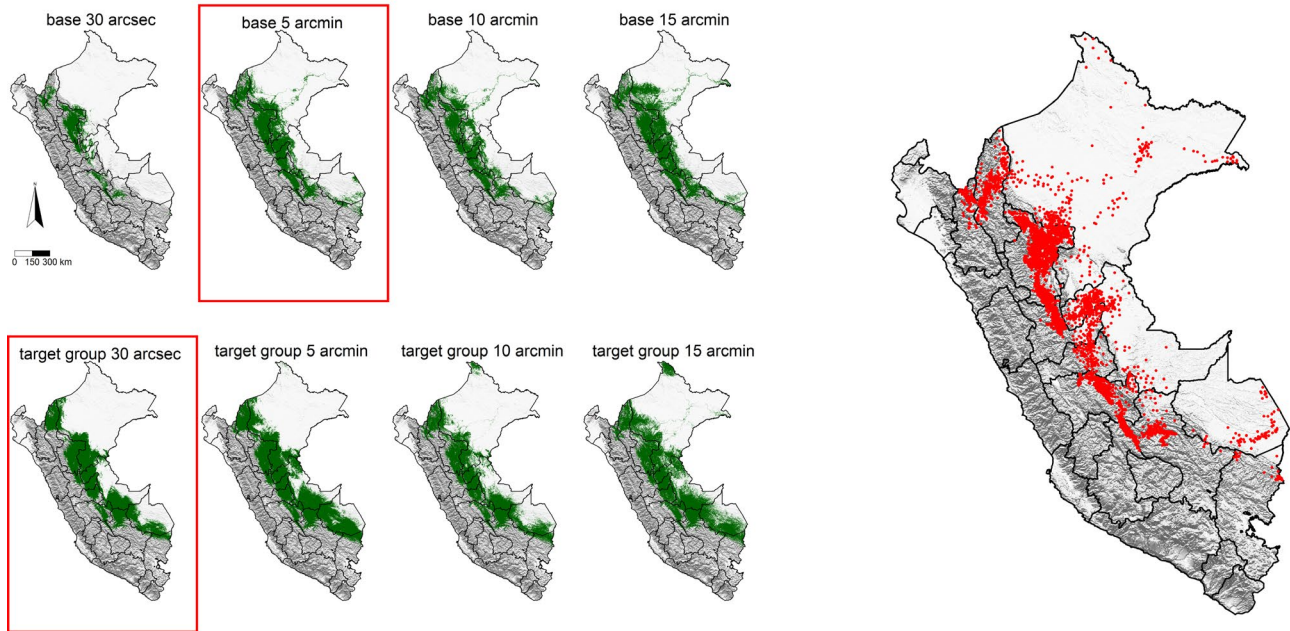
(b) Wild cacao

FIGURE 1 Prediction of suitable habitat for all the ensemble models constructed with different pseudo-absence points selection approaches and resolutions of spatial filtering for a) cultivated and b) wild cacao in Peru. The best-performing models for each of the pseudo-absence point selection approaches are in red boxes: the base model at 5 arcmin filtering and the target group model at 30 arcsec filtering for cultivated cacao, and the base model at 30 arcsec filtering and the target group model at 30 arcsec filtering for wild cacao. Green areas represent the suitable area for cacao predicted by each model in the left maps. Presence points used for the modelling are shown in red in the right maps

we retrieved all presence points within the geographical extent of Peru (66.8–81.6°W; 0.4–19.3°S), also including points from neighbouring countries Colombia, Ecuador and Brazil within this extent (Figure 1b). Cultivated cacao in Peru spans at least seven of the 10 genetic groups currently described in cacao (Motamayor et al., 2008)

and includes also many introduced cacao varieties. Therefore, in our dataset, we considered both national and international cultivated cacao varieties present in Peru (Table S1 in Appendix S1). For cultivated cacao, presence points from the coastal regions and other sites with <600 mm of annual precipitation were omitted as cacao

can only be grown in these regions under irrigation, which would confound the identification of naturally suitable areas. Accordingly, there are no records of wild cacao in these regions. The compiled presence points comprised 19,685 points for cultivated and 1,182 for wild cacao.

2.2 | Environmental variables

The explanatory variables included 34 climate, soil and terrain variables commonly used in habitat suitability modelling (Table S2 in Appendix S1). These variables were selected because climate, soil and terrain are the main factors influencing species distributions. Collinear variables under present conditions were removed based on stepwise calculations of variance inflation factors (VIF), retaining only variables with VIFs <10. As correlations between variables can change both in strength and direction in the future (Braunisch et al., 2013; Dormann et al., 2013), we also assessed collinearity under future conditions and retained all variables resulting from the stepwise VIF procedure in at least one time horizon-by-emission scenario combination (see below). Although this resulted in a final set of variables of which some had a VIF >10, this does not outweigh the risk of excluding variables with diverging collinearity patterns under future conditions, which may lead to misleading predictions of future habitat suitability (Braunisch et al., 2013). Multicollinearity analyses were based on the values of the explanatory variables at the locations of all presence and pseudo-absence points used to calibrate each of the ensemble models (see below).

2.3 | Habitat suitability modelling

We generated multiple ensemble models (see below) for both cultivated and wild cacao using different approaches to select pseudo-absence points, and different resolutions of spatial filtering. We selected the pseudo-absence points of both approaches from the area enclosed by a convex hull polygon around all presence points extended with a buffer corresponding to 10% of the polygon's largest axis (Acevedo et al., 2012). We further restricted the area to sites below 2500 masl (hereafter called the "convex hull area") to improve the ability of the suitability models to distinguish between presence and absence at lower altitudes.

We used two approaches for the selection of pseudo-absence points: random selection (from hereon called the "base approach") and target group selection ("target group approach"). We separately selected pseudo-absence points for the presence–absence algorithms in our ensembles (see below) and background points for Maxent algorithm. In the base approach, background points and pseudo-absences were randomly selected from the convex hull area. In the target group approach, pseudo-absence and background points were selected from the convex hull area according to the method outlined by Phillips et al., (2009) and Mateo et al., (2010), which involves the selection of points from grid cells with presences of species belonging to a similar group as the target species. These locations are expected to reflect a similar sampling bias compared

with the target species, thus reducing the effects of spatially biased presence points on model calibration (Phillips et al., 2009). The target group grid to select pseudo-absence points was based on the farm locations in the Peruvian national agricultural census of 2012 for cultivated cacao and on the presence points of all tree and shrub species occurring in the national checklists of Peru, Ecuador and Colombia for wild cacao (see Text S2 in Appendix S1 for more details).

Both pseudo-absence point selection approaches for cultivated and wild cacao were crossed with different resolutions of spatial filtering. Spatial filtering is used to reduce the effects of spatially biased presence points on model calibration (Kramer-Schadt et al., 2013) and was implemented by randomly retaining one presence point per grid cell for each of the selected resolutions of spatial filtering. We carried out spatial filtering at four resolutions, ranging from 30 arcsec (ca. 0.9 km at the equator) to 15 arcmin (ca. 27 km). For the target group approach, we also filtered the pseudo-absence points to the respective resolutions of the models to reflect the spatial bias of the filtered presences. For the base approach, we kept the original resolution of pseudo-absence points as randomly selected points should not show spatial sorting bias. As such, we calibrated eight ensemble models for cultivated cacao and eight ensemble models for wild cacao (two pseudo-absence points selection approaches combined with four resolutions of spatial filtering).

Suitability modelling was performed using ensembles composed of up to 19 modelling algorithms, using the *BiodiversityR* package for R (Kindt, 2018). The list of the 19 modelling algorithms is presented in Table S3 in Appendix S1. To carry out an initial selection of the algorithms to include in the ensemble and remove spatial sorting bias (Hijmans, 2012), we performed a 20-fold cross-validation for each algorithm using the presence and pseudo-absence points of each of the combinations of pseudo-absence selection and spatial filtering, calculated the calibrated Area Under the receiver-operator Curve (cAUC) for each of the cross-validation folds and retained the algorithms with cAUC values significantly higher than the null model according to Mann–Whitney tests, following Thomas et al., (2014).

The retained algorithms in the ensemble models were cross-validated for 10 folds using spatial blocks with the *blockCV* package for R (Valavi et al., 2019), following Fremout et al., (2020). With the spatial block cross-validation, presence and pseudo-absence points were partitioned in training and testing points using a set of spatial blocks, with each fold consisting of one or more 200 km wide squared blocks. AUC values cross-validated with spatial blocks (cvAUC) provide a better measure of model transferability, which is crucial when projecting species distributions to future climates (Wenger & Olden, 2012). The weights of the algorithms in each of the ensemble models were optimized using the *ensemble.tune* function in *BiodiversityR* (Kindt, 2018). The ensemble models were then calibrated again using all presence and pseudo-absence points and the previously optimized weights. For each of the ensemble models, we calculated four evaluation metrics (cvAUC, False Negative Rate (FNR), True Skill Statistic (TSS) and Symmetric Extremal Dependence Index (SEDI)) (Wunderlich et al., 2019). The best model of each approach was selected on the basis of having the highest cvAUC, TSS and SEDI, and lowest FNR.

(a) Cultivated cacao
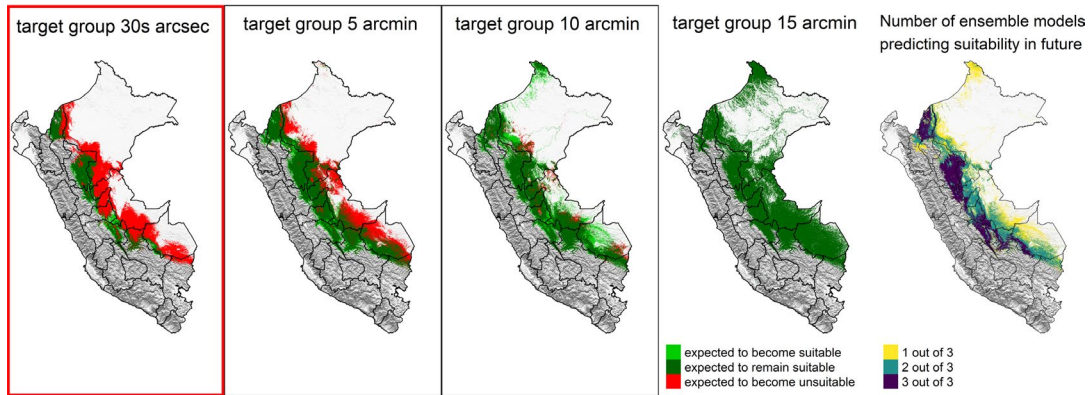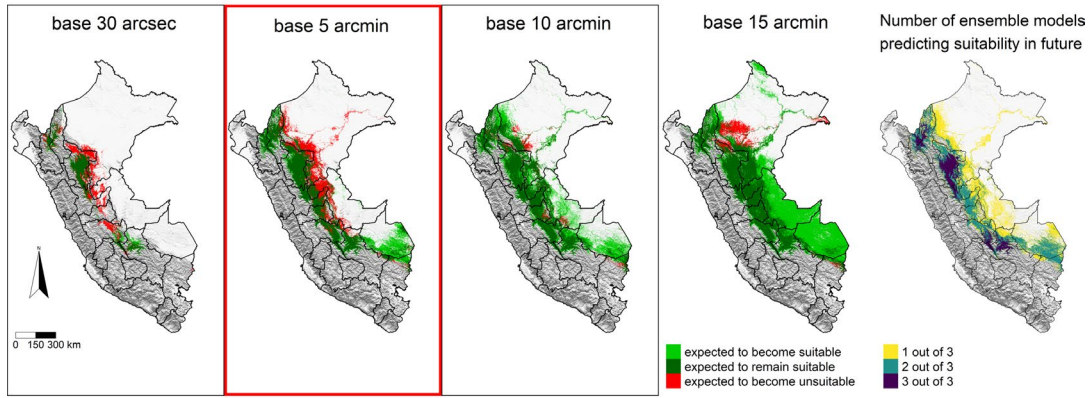
(b) Wild cacao
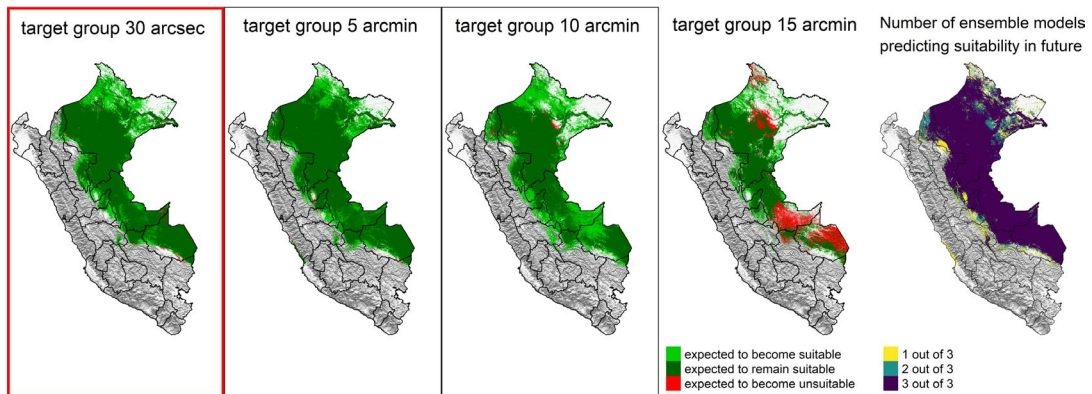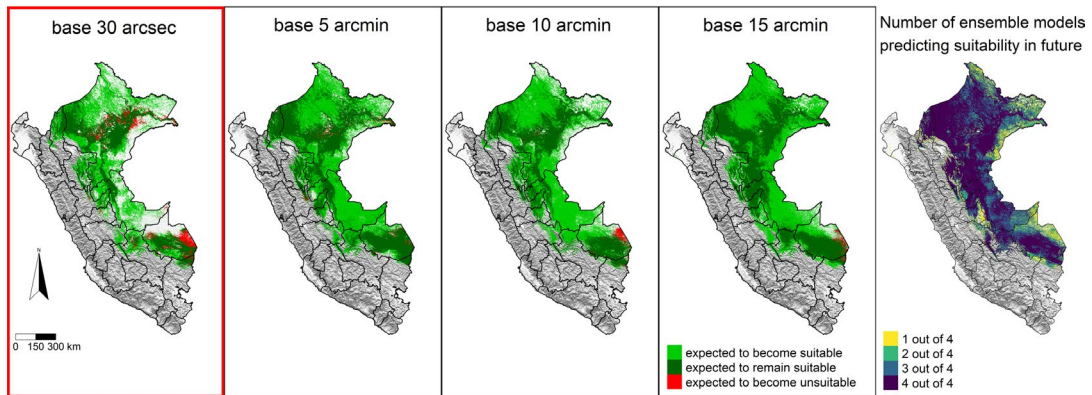
The importance of the environmental variables retained in the four best-performing ensemble models was calculated using the *biomod2* package for R (Thuiller, 2003), by calculating the Pearson's correlation between the original suitability predictions and the suitability predictions obtained with a shuffled dataset. Response curves of each of the environmental variables were produced using the *evaluation.strip.plot* function in the *BiodiversityR* package.

Future projections of cacao suitability were made for three time horizons (2030s, average for 2020–2049; 2050s, average for 2040–2069; 2070s, average for 2060–2089) and two Representative Concentration Pathways (RCP4.5 and RCP8.5). RCP4.5 corresponds to the emission scenario in which all countries would comply with their Intended Nationally Determined Contributions under the Paris Agreement (Tribett et al., 2017), whereas the RCP8.5 represents a worse-case scenario. Future climatic conditions were characterized using multiple downscaled General Circulation Models (GCMs) from the CCAFS Climate Data Portal (Ramirez Villegas & Jarvis, 2010). We compared the future predictions obtained using all the GCMs available on this platform (31 GCMs for RCP4.5 and 33 GCMs for RCP8.5) with a subset of five well-performing and dissimilar GCMs (CESM1(CAM5), GFDL-CM3, HADGEM2-ES, MIROC5 and MPI-ESM-LR). This subset was selected following Fremout et al., (2020) (Text S3 in Appendix S1).

Suitability maps for present and future climates were converted in presence–absence maps using the maximum training sensitivity plus specificity threshold (Liu et al., 2005). Suitable conditions under future climatic conditions were predicted in grid cells where at least 75% of the GCMs coincide. For each of the six time horizon-by-RCP combinations, we defined novel climates (i.e., climatic conditions outside the range between the 5% and 95% percentile of the conditions used to calibrate the distribution models) using the *ensemble.novel* function in *BiodiversityR*. For simplicity, the figures in the manuscript focus on the projections for 2050s and RCP4.5, but we present the main findings for all the time horizon-by-RCP combinations in the text and include figures for all combinations in the appendix.

2.4 | Ecogeographical zones and niche overlap analysis

To assess changes in growth conditions within the suitable distribution range of cultivated and wild cacao under climate change, we subdivided the current cacao distribution in homogeneous ecogeographical zones (Parra-Quijano et al., 2012) and projected

these zones to future climates, following Fremout et al., (2021). We determined these zones using the CLARA (Clustering Large Applications) method implemented in the *cluster* package for R (Maechler et al., 2019). As input data, we used the first 10 principal components from a principal component analysis using values for all the 34 environmental variables mentioned above within the total current distribution range of cacao (both cultivated and wild), which together explained more than 95% of variance. The potential number of clusters (i.e., ecogeographical zones) was varied between five and 25, and the optimal number was determined using the cluster silhouette criterion through the *optCluster* package for R (Sekula et al., 2020), resulting in nine zones. Variables were normalized prior to clustering. For future projections, each grid cell was assigned to the closest cluster in ordination space based on the values of the principal components resulting from the environmental variables under future climate scenarios, using the *cl_predict* function from the *clue* package for R (Hornik & Böhm, 2020).

To investigate whether cultivated and wild cacao occupy different environmental niches, we performed a niche overlap analysis and niche hypervolume analysis. We calculated niche overlap, niche equivalency and niche similarity following the methodology described by Broennimann et al., (2012) and implemented in the *ecospat* package for R (Broennimann et al., 2012). We calculated the niche hypervolume occupied by cultivated and wild cacao and their intersection using the *hypervolume* package for R (Blonder et al., 2014). Further methodological details are presented in Text S4 in Appendix S1.

2.5 | Outlier analysis for climate change-tolerant genotypes

We carried out an outlier analysis of the environmental conditions within the modelled distribution ranges to identify the areas where climate change-tolerant genotypes are potentially present. For this analysis, we selected twelve variables predicted to be important for climate change adaptation of cacao cultivation based on variable importance, suitability response curves and the expected direction of climate change (Table 2, Fig. S6, Fig. S7 and Fig. S8 in Appendix S1). These variables include flooding risk (tolerance to flooding), altitude (tolerance to occasional chills) and ten climatic variables: maximum temperature of warmest month, mean temperature of driest quarter, annual precipitation, precipitation of wettest quarter, precipitation

FIGURE 2 Future distribution of suitable habitat according to all the ensemble models constructed with different absence point selection approaches and resolutions of spatial filtering for a) cultivated and b) wild cacao in Peru. Future projections show all areas identified as suitable by at least 75% of the subset of five well-performing and dissimilar General Circulation Models (GCMs) for the Representative Concentration pathway (RCP) 4.5 and the 2050s time period. Ensemble models with sufficient predictive accuracy (cvAUC >0.75) used for estimate future uncertainty are in black boxes while the best models for each approach are in red boxes: the base model at 5 arcmin filtering and the target group model at 30 arcsec filtering for cultivated cacao, and the base model at 30 arcsec filtering and the target group model at 30 arcsec filtering for wild cacao. The first four maps present the areas expected to become suitable (light green), the area expected to remain suitable (dark green) and the area expected to become unsuitable (red) as predicted by each of the four models. As a summary, the consensus maps on the right present the number of ensemble models predicting suitability (from 1 model in yellow up to 4 models in dark purple), including only the ensemble models with sufficient predictive accuracy

of driest quarter, annual potential evapotranspiration, aridity (tolerance to heat and drought), mean diurnal range, temperature seasonality and precipitation seasonality (tolerance to altered diurnal and seasonal variation). The flooding risk map was retrieved from the National Service of Meteorology and Hydrology of Peru (SENAMHI, Spanish acronym) (Aybar et al., 2019). Altitude was included among the selected in the view of potential expansion of cacao at higher elevations where occasional chills can occur. Apart from flooding risk and altitude, among the remaining ten climatic variables, annual precipitation, precipitation of wettest quarter and aridity were colinear with other variables and not retained by the models and therefore do not have a response curve. Nevertheless, they were used in the outlier analysis because annual precipitation is one of the most important factors influencing yield (Zuidema et al., 2005), precipitation of the wettest quarter is an important predictor of flooding risk, to which cacao can be sensitive (De Almeida & Valle, 2007; Wood & Lass, 2008), and aridity is a fundamental indicator for climate change assessments (Berg et al., 2016; Fu & Feng, 2014).

For each variable, we calculated the outlier threshold as the 1% percentile for the lower outliers and the 99% percentile for the upper outliers. The outliers were calculated based on all grid cell values within the total cacao distribution range (both cultivated and wild). To maximize the chance of including areas under the most extreme environmental conditions at the margins of habitat distribution, we defined the distribution range for wild cacao using the minimum training presence (i.e. the lowest predicted suitability value at the training presence points) as threshold for suitability. Coastal regions where wild cacao does not occur were excluded from the obtained distribution for wild cacao by excluding the respective terrestrial ecoregions. The areas where climate change-tolerant genotypes are potentially present were identified as those with environmental variable scores above or below the respective outlier thresholds. Finally, we maximized the probability of actual presence by overlaying these areas with 5 km buffers around cacao presence points.

3 | RESULTS

The calibrated suitability models resulted in different suitability distributions, depending on the pseudo-absence point selection approach and resolution of spatial filtering (Figure 1). Table S4 in Appendix S1 shows the number of presence and pseudo-absence points, modelling algorithms and environmental variables used in the different ensemble models. In general, models generated with the target group approach and at lower resolutions of spatial filtering resulted in broader and less overfit distributions compared with models constructed using the base approach and higher resolutions of spatial filtering (Figure 1). According to the evaluation metrics (Fig. S1 in Appendix S1), the best-performing models for each of the pseudo-absence points selection approaches were the base model at 5 arcmin filtering and the target group model at 30 arcsec filtering for cultivated cacao, and the base model at 30 arcsec filtering and the target group model at 30 arcsec filtering for wild cacao. For

cultivated cacao, we were interested in the model that provides the closest fit to the current range of cultivated cacao (i.e. base model at 5 arcmin filtering) to obtain more reliable indications of climate change impacts in currently cultivated areas. By contrast, for wild cacao the target group model at 30 arcsec provides a more realistic approximation of the present distribution of wild cacao compared with the base approach models which tended to overfit the spatially biased presence data. The selected models for cultivated and wild cacao presented distinct suitability distributions, with cultivated cacao extending along the Andes foothills and wild cacao covering most of the Peruvian Amazon (Figure 1). In both models, the coastal regions were not predicted as suitable.

To estimate the uncertainty of future predictions related to model selection, we generated projections for all the ensemble models and built consensus maps including only the models with sufficient predictive accuracy (cvAUC >0.75) (Figure 2). As the predictions of the four best-performing ensemble models using the full set of GCMs (Figure 3 and S2) were similar to the predictions obtained with the subset of five well-performing and dissimilar GCMs (Fig. S3 in Appendix S1), we used only the subset of five GCMs for these projections. For cultivated cacao, spatial filtering strongly influenced the future projections (Figure 2a), with forecasts becoming more optimistic at decreasing resolution of spatial filtering, regardless of the pseudo-absence selection method used. For wild cacao, future projections were more consistent among different ensemble models (Figure 2b).

The future projections of the four best-performing models showed different trends between cultivated and wild cacao. For cultivated cacao, future projections predict a contraction of suitable areas in the northern and central Amazon, and an expansion into the southern Amazon of Madre de Dios as well as towards higher elevations along the Andes (Figure 3 and S2 in Appendix S1). A comparison of current and future suitability models showed that the 99% quantile for altitude shifts from 1,700 masl under present climate conditions up to 2,100 masl for the expected future distribution range (Fig. S4 in Appendix S1). However, according to at least 75% of the GCMs, the predicted expansion areas are smaller than the contraction areas, resulting in a net decrease in suitable area (Table 1 and Table S5). Similar trends are expected under the worst-case (RCP8.5) and more optimistic emission scenario (RCP4.5), and different time horizons (Fig S5). For wild cacao, most of the current distribution range is foreseen to remain suitable by at least 75% of GCMs and is predicted to further expand towards the Amazon region and the Andes (Figure 3 and Figure S2). Higher gains in suitable area of wild cacao are predicted at more distant time horizons and under the worst-case emission scenario (RCP8.5) (Table 1 and Table S5).

The single best suitability models for cultivated (base model at 5 arcmin filtering) and wild cacao (target group model at 30 arcsec filtering) retained similar environmental variables, although these variables showed different importance scores (Fig. S6 in Appendix S1) and different trends in the response curves (Table 2, Fig. S7 in Appendix S1). The niche overlap and niche hypervolume analyses showed that cultivated and wild cacao occupy different

(a) Cultivated cacao

Change of future suitability respect to present for at least 50% GCMs
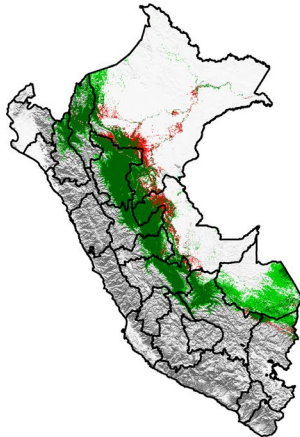
Change of future suitability respect to present for at least 75% GCMs

Change of future suitability respect to present for at least 90% GCMs

N
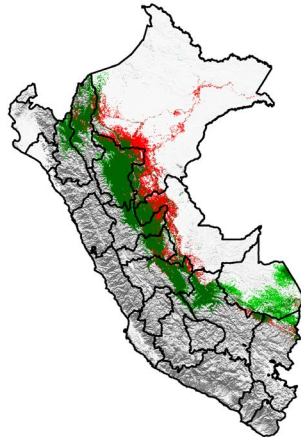
0 150 300 km

expected to become suitable
expected to remain suitable
expected to become unsuitable

(b) Wild cacao

Change of future suitability respect to present for at least 50% GCMs

Change of future suitability respect to present for at least 75% GCMs

Change of future suitability respect to present for at least 90% GCMs

N

0 150 300 km

expected to become suitable
expected to remain suitable
expected to become unsuitable

FIGURE 3 Future distribution of suitable habitat for a) cultivated and b) wild cacao in Peru showing all areas identified as suitable by at least 50%, 75% and 90% of 31 different future General Circulation Models (GCMs). The maps present the areas expected to become suitable in light green, the area expected to remain suitable in dark green and the area expe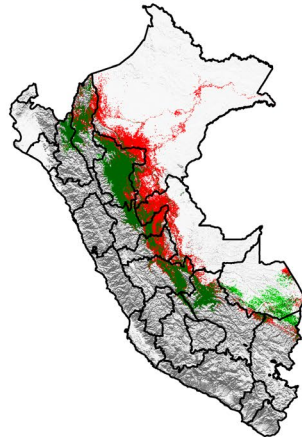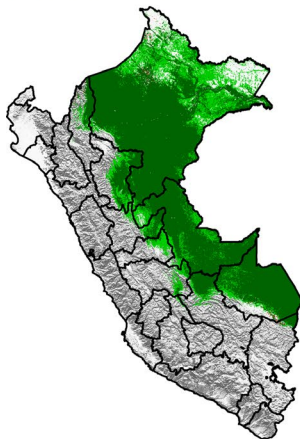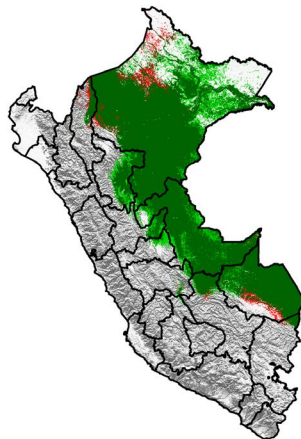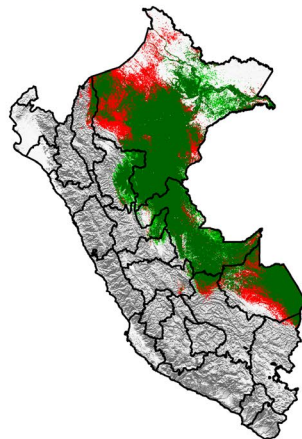cted to become unsuitable in red. Here, we present the maps for the selected ensemble models (i.e. the base model at 5 arcmin filtering for cultivated cacao and the target group model at 30 arcsec filtering for wild cacao), and for the Representative Concentration Pathway (RCP) 4.5 and the 2050s time horizon, but similar results were found for the projections with combinations of other models, emission scenarios and time horizons (Figs. S2 and S5, Table 1 and S5)

environmental niches, with cultivated cacao occurring in areas with higher mean diurnal temperature range, lower mean temperature of wettest quarter and higher precipitation of the warmest quarter compared with wild cacao (Text S4).

In the area where cacao is currently cultivated, temperatures and potential evapotranspiration are predicted to increase across the whole distribution range, while precipitation changes show varying patterns depending on the season and region (Fig. S8 in Appendix S1). For cultivated cacao, expansion of suitable range

seems to occur in areas that will have higher precipitation of wettest month/quarter and coldest quarter, and similar precipitation of driest month/quarter and warmest quarter compared with current precipitation regimes in currently suitable areas (Fig. S8 in Appendix S1). Similar changes in climate are expected within the distribution range of wild cacao (Fig. S9 in Appendix S1).

Ecogeographical zones identified within the cacao distribution range are expected to change 8%–16% in area (Fig. S10, Table S6 in Appendix S1), meaning that local climatic conditions are expected to

TABLE 1 Changes in suitable areas for cultivated and wild cacao in Peru for two Representative Concentration Pathways (RCPs) and three time horizons. Total suitable areas and area changes were calculated by predicting suitable conditions in grid cells where at least 75% of the 31/33 General Circulation Models (GCMs) predict suitable conditions. Here, we present the results for the selected ensemble models (i.e. the base model at 5 arcmin filtering for cultivated cacao and the target group model at 30 arcsec filtering for wild cacao), while the results for the target group model at 30 arcsec filtering for cultivated cacao and the base model at 30 arcsec filtering for wild cacao are presented in Table S5 in Appendix S1

		CULTIVATED CACAO			WILD CACAO		
RCP	**Time horizon**	**Contraction area**	**Expansion area**	**Net area change**	**Contraction area**	**Expansion area**	**Net area change**
RCP4.5	2030s	−26%	+17%	−9%	−7%	+12%	5%
	2050s	−29%	+18%	−11%	−4%	+17%	13%
	2070s	−28%	+19%	−9%	−4%	+18%	15%
RCP8.5	2030s	−26%	+20%	−6%	−3%	+15%	13%
	2050s	−27%	+23%	−4%	−1%	+29%	28%
	2070s	−31%	+25%	−6%	−0%	+39%	39%

change even in some areas that are predicted to remain suitable. No novel climatic conditions are expected to appear within the present nor future distribution range of cultivated and wild cacao according to most of the GCMs (Fig. S11 in Appendix S1), except for small areas in Puno and Ucayali for the base model of wild cacao (Fig. S11c in Appendix S1).

According to the selected variables, potential climate change-tolerant populations are spread across different regions around Peru and in many cases occur along the margins of the cacao distribution range (Figure 4). Some areas present multiple environmental extremes, for instance the areas in the departments of Cuzco and Cajamarca.

4 | DISCUSSION

In this study, we modelled the current and future distribution of cultivated and wild cacao in Peru by comparing multiple ensemble models constructed using different pseudo-absence selection approaches and resolutions of spatial filtering. Both the procedures had pronounced impacts on model predictions (Figure 1). In general, the target group approach and spatial filtering at lower resolutions reduced overfitting by reducing spatial sample bias (Phillips et al., 2009). However, spatial filtering strongly impacted future predictions for cultivated cacao, ranging from net negative impacts at higher spatial resolution of filtering to net positive impacts at lower resolution. And this in spite of the fact that models calibrated with data filtered at different spatial resolutions had a similar accuracy in terms of cvAUC and produced similar distribution ranges under current climate conditions.

Two lessons can be learned from this. First, that model selection must be tailored to the characteristics of available data, the modelled species and the intended use of the suitability models (Derville et al., 2018; Segurado & Araujo, 2004). In our case, for cultivated cacao we had very detailed agricultural census data and were interested in the current (instead of potential) range of cultivated cacao to obtain more reliable indications of impacts of climate change in currently cultivated areas. By contrast, for wild cacao our data showed

strong spatial sampling bias and we were interested in a model that best predicts the current realized distribution also in poorly sampled areas. Hence, the random pseudo-absence point selection worked better for cultivated cacao, while the target group selection worked better for wild cacao. A second lesson is that selecting the best models for future climate projections based on performance under current climate conditions alone may be misleading as alternative models with similar accuracy can result into variable future projections (Araújo & New, 2007; Thuiller et al., 2019). Therefore, testing projections for multiple models is required to estimate the uncertainty of future predictions (Araújo & New, 2007; Thuiller et al., 2019).

Our results show that cultivated and wild cacao present distinct suitability distributions (Figure 1), in accordance with the fact they occupy different environmental niches (Text S4 in Appendix S1). This is in line with other suitability modelling studies comparing the distributions of cultivated and wild populations of the same species (d'Eeckenbrugge & Lacape, 2014; Eshetae et al., 2019; A. J. Miller & Knouft, 2006). The best-performing model obtained for cultivated cacao presents a realistic representation of where cacao is currently grown in Peru (Figure 1a). For cultivated cacao, many parts of Peruvian Amazon are probably currently suitable, as evidenced by cacao cultivation by indigenous communities and small-scale plantations (Levis et al., 2017; R. P. Miller & Nair, 2006), but cacao is not grown there at commercial scale. As expected, none of the models of cultivated cacao predicted suitability in the coastal regions where cacao is unable to grow without irrigation. The best-performing model for wild cacao represents a realistic distribution of the current suitable habitat of wild cacao (Figure 1b). However, even though the target group suitability models for wild cacao minimized overfitting, the best-performing model still omitted some areas that may be suitable in the northern Amazon region and assigned higher suitability values to the more accessible areas along rivers. This is most likely due to the strong spatial sample bias, with a near complete lack of collection missions in inaccessible areas located at longer distances from rivers.

TABLE 2 Overview of general trends in response curves of cultivated and wild cacao suitability (Y-axis) to different climate variables (X-axis) and trends of climate variables under climate change. Red trend lines indicate whether the variable has positive, negative, flat, flattening or optimum curve (exact response curves are given in Fig S7). Black arrows indicate whether the variable is expected to increase, decrease, or increase and decrease depending on the areas in the future in the distribution range of cacao in Peru (shown in Fig S8). The climatic variables included in the outlier analysis to identify climate-tolerant genotypes are indicated with an "a". These response curves and future trends of climatic variables were used to define the ten climatic variables used in the outlier analysis, in addition to altitude and flooding risks. The selected climatic variables are: maximum temperature of warmest month, mean temperature of driest quarter, annual precipitation, precipitation of wettest quarter, precipitation of driest quarter, annual potential evapotranspiration, aridity, mean diurnal range, temperature seasonality, precipitation seasonality. Among these ten variables, annual precipitation, precipitation of wettest quarter and aridity were colinear with other variables and not retained by the models, and therefore do not have a response curve. Nevertheless, they were used in the outlier analysis because annual precipitation is one of the most important factor defining yield (Zuidema et al., 2005), precipitation of the wettest quarter is the main predictor of flooding risk which cacao is sensitive to (Almeida & Valle, 2007; Wood & Lass, 2008), and aridity is a fundamental indicator for climate change assessments (Berg et al., 2016; Fu & Feng, 2014)

Variable	Response curve of cultivated cacao	Response curve of wild cacao	Future trend
Mean Diurnal Range[a]			↓↑
Isothermality			↓
Temperature Seasonality[a]			↑
Max Temperature of Warmest Month[a]			↑
Temperature Annual Range			↓↑
Mean Temperature of Wettest Quarter			↑
Mean Temperature of Driest Quarter[a]			↑
Precipitation of Wettest Month			↑
Precipitation of Driest Month			↓↑
Precipitation Seasonality[a]			↓↑
Precipitation of Wettest Quarter[a]			↑
Precipitation of Warmest Quarter			↓
Precipitation of Coldest Quarter			↑
Annual Potential Evapotranspiration[a]			↑

[a]Climatic variables included in the outlier analysis to identify areas where climate tolerant populations are potentially present.

Future projections yielded different results between cultivated and wild cacao. For cultivated cacao, projections foresee a contraction of the suitable range and shift at higher elevation which is in line with other studies in West Africa (Läderach et al., 2013; Schroth et al., 2016) and Central America (Eitzinger et al., 2015; de Sousa et al., 2019), although there is some uncertainty among projections from different ensemble models (Figure 3). Because cacao cultivation in Peru has been expanding at an annual rate of 9.1% in recent years (MINAGRI, 2018), new expansion efforts

should be directed towards areas that will remain or will become suitable into the future rather than the predicted contraction areas in the northern and central Amazon (Figure 3). However, some of the new suitable areas in southern Amazon are currently covered by forests and it is paramount that potential expansion of cacao cultivation is not implemented at the expense of deforestation or encroachment into protected areas. By contrast, our projections consistently foresee a positive future for wild cacao in Peru as the current suitable area is expected to largely remain suitable and to
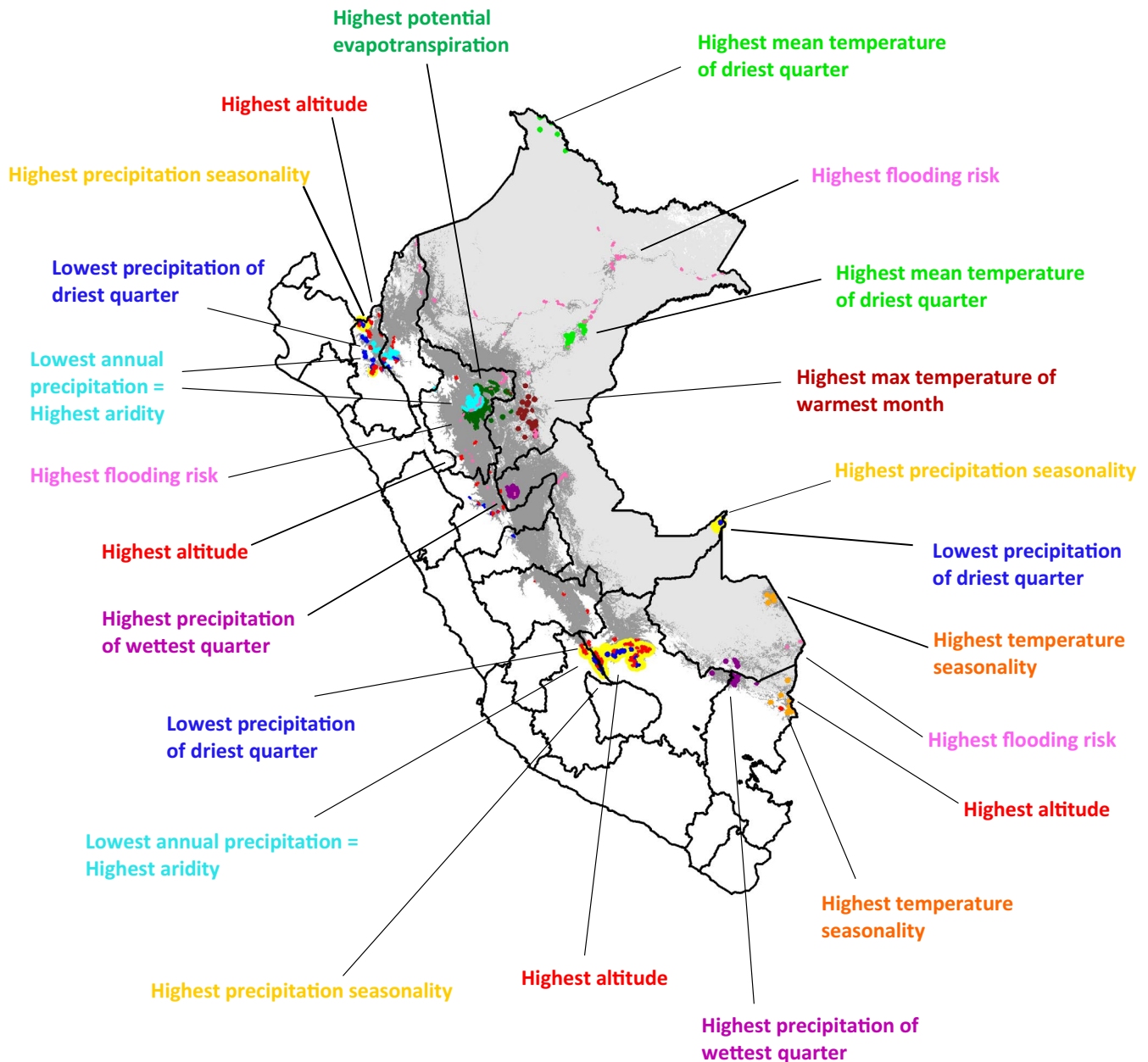
FIGURE 4 Areas where cacao occurs under extreme environmental conditions within Peru for the twelve selected variables, as identified by the outlier analysis. The area in dark grey represents the distribution range of cultivated cacao (base model at 5 arcmin filtering) while the area in light grey represents the distribution range of wild cacao (target group model at 30 arcsec filtering). The distribution range for wild cacao was defined using the minimum training presence as threshold for suitability, excluding the terrestrial ecoregions (Olson & Dinerstein, 2002) without presence of wild cacao. The areas shown are identified by overlaying the areas with extreme environmental conditions with 5km buffers around known cacao presence points.

further expand in the future. As mentioned above, predicted expansion areas in the northern Amazon region are possibly already suitable for wild cacao, but the Andes foothills represents a newly suitable area for wild cacao.

Different responses of cultivated and wild cacao to climate change might be explained by the fact the two groups are partly adapted to and/or grown under different climatic conditions, as shown by the distinct suitable distributions and environmental niches (Figure 1, Text S4 in Appendix S1), and by the different variable importance scores and response curves (Table 2, Fig. S6 and

Fig. S7 in Appendix S1). Different importance scores and mean values in the environmental variables have also been found in other modelling studies comparing cultivated and wild populations of the same species (Eshetae et al., 2019; Galluzzi et al., 2015; A. J. Miller & Knouft, 2006). During domestication, humans exert artificial selection on a species, rendering the cultivated populations distinct from the wild progenitors and eventually resulting in slightly different environmental adaptations. In our study, differences in future projections can be partly explained by the fact that for wild cacao the response curves between climate variables and

habitat suitability more often follow the same trend as the direction of climate change than for cultivated cacao. For example, our results (Table 2, Fig. S7 and Fig. S8 in Appendix S1) suggest that cultivated cacao is vulnerable to expected future decreases in diurnal and increases in annual temperature ranges, and increases in temperature and evapotranspiration, whereas wild cacao might be better adapted to those conditions. However, further research is needed to understand to what extent the response curves reflect the current distribution ranges (realized niche) of cultivated and wild cacao, as compared to the true breath of environmental conditions under which cacao populations could thrive (fundamental niche). For example, evidence from West Africa and Malaysia suggest that cacao can be cultivated at higher potential evapotranspiration and temperatures (maximum daily temperatures up to 44°C) than currently observed in Peru, as long as water availability is not limiting (Lahive et al., 2019; Medina & Laliberte, 2017; Wood & Lass, 2008).

The search for climate change-tolerant cacao genotypes will become increasing important in the areas where cultivated cacao is expected to loose suitability and also in areas predicted to remain suitable but where a shift in ecogeographical zones is expected. Due to the lack of sufficient reliable data on individual varieties, we modelled cultivated cacao as a single entity assuming uniform adaptation to environmental conditions across the distribution area. In reality, different varieties may vary in local adaptation to the environment (Hällfors et al., 2016; Smith et al., 2017). This is especially true in Peru where a large diversity of native cacao varieties exists (Table S1). As the environmental niches of different varieties are likely to be subsets of the environmental niche that we have modelled, even in areas that are predicted to remain suitable, individual varieties may not be adapted to the predicted changes and may require to be replaced with more tolerant ones. Lowland Amazonian cacao populations have great potential for facilitating the adaptation of cacao cultivation to climate change, considering that wild cacao seems to be more resistant to climate change. Furthermore, as these putatively wild populations in fact represent a gradient of wild over cultivated to incipiently domesticated cacao genetic resources which are the result of thousands of years of management by indigenous societies (Clement et al., 2020; Levis et al., 2017; Thomas, 2017), they may hold many more traits of human interest than just resistance to extreme climate conditions.

We identified areas where cacao populations are expected to be present that are tolerant to (a) drought, heat and increased seasonality, because they are fundamental for cacao suitability and are expected to change the most in the future, (b) occasional chills, because future projections predicted elevational shifts, and (c) flooding, because cacao is sensitive to flooding (A. De Almeida & Valle, 2007; Wood & Lass, 2008). For climate adaptation purposes, it may be advantageous to select climate change genotypes from areas where multiple extremes overlap, on the premise that these may be tolerant to multiple stress factors simultaneously. For example,

several areas in Cajamarca and Cuzco combine potential drought tolerance with the tolerance to growth conditions at high elevations.

5 | CONCLUSIONS AND FUTURE PROSPECTS

In this study, we modelled the distribution of suitable habitat of cultivated and wild cacao in Peru by comparing multiple ensemble models constructed using different pseudo-absence selection approaches and resolutions of spatial filtering. We found that model selection significantly impacted both present and future modelling results. This stresses the importance of tailoring the modelling approach to the objectives and characteristics of the target species and projecting multiple models to future climatic conditions, to better assess uncertainty of future projections, which should be considered in decision-making related to climate change policies (Araújo & New, 2007; Thuiller et al., 2019).

Overall, our models foresee a contraction of suitable area for cultivated cacao while predicting a more positive future for wild cacao in Peru. Because suitability is not necessarily correlated with yield quantity or quality (Ramirez-Villegas et al., 2013), future studies should asses the expected impacts of climate change on cacao yields, also including indirect effects such as altered incidence of pests and diseases (Gateau-Rey et al., 2018).

Tolerant genotypes will be required to facilitate the adaptation of cacao cultivation under climate change in Peru. To this aim, populations of native cacao in the identified areas will be target of planned collection missions. The resistance of these genotypes to extremes in growth conditions will be verified in greenhouse and climate chamber experiments. The most promising genotypes will then be established in clonal gardens for the supply of the propagation material. In the longer term, these genotypes should be included in breeding programmes aimed not only at enhancing climate change tolerance but also yield quantity and quality, pest resistance and low heavy metal accumulation.

Finally, the climate change-tolerant genotypes identified in this study could also benefit other cacao producing countries where climate change is expected to cause a contraction of the suitable area for cacao cultivation (Eitzinger et al., 2015; Läderach et al., 2013; Schroth et al., 2016; de Sousa et al., 2019). Especially genotypes tolerant to high temperatures and low precipitation/ high evapotranspiration have the potential to increase the climate change resilience of cacao production systems in these countries, as these are the factors expected to drive contraction of suitable area for cacao cultivation in future (Läderach et al., 2013; Schroth et al., 2016).

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ddi.13294.

DATA AVAILABILITY STATEMENT

The dataset with the presence points used for modelling cultivated and wild cacao is available in Appendix S2.

ORCID

Viviana Ceccarelli https://orcid.org/0000-0003-2160-9483
Tobias Fremout https://orcid.org/0000-0002-0812-3027
Sphyros Lastra https://orcid.org/0000-0001-9017-0852
Enrique Arévalo-Gardini https://orcid.org/0000-0002-1725-6788
Evert Thomas https://orcid.org/0000-0002-7838-6228

REFERENCES

Acevedo, P., Jiménez-Valverde, A., Lobo, J. M., & Real, R. (2012). Delimiting the geographical background in species distribution modelling. *Journal of Biogeography*, *39*(8), 1383–1390. https://doi.org/10.1111/j.1365-2699.2012.02713.x

Apshara, S. E., Hebbar, K. B., Mathias, T. G., & Séverin, A. (2016). Morpho-physiological criteria for assessment of two month old cocoa (Theobroma cacao L.) genotypes for drought tolerance. *Indian Journal of Plant Physiology*, *21*(1), 23–30. https://doi.org/10.1007/s40502-015-0195-y

Araque, O., Jaimez, R. E., Tezara, W., Coronel, I., Urich, R., & Espinoza, W. (2012). Comparative photosynthesis, water relations, growth and survival rates in juvenile Criollo cacao cultivars (Theobroma cacao) during dry and wet seasons. *Experimental Agriculture*, *48*(4), 513–522. https://doi.org/10.1017/S0014479712000427

Araújo, M. B. & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, *22*(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., & Felipe-Obando, O. (2019). Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. *Hydrological Sciences Journal*, 1–16, https://doi.org/10.1080/02626667.2019.1649411

Berg, A., Findell, K., Lintner, B., Giannini, A., Seneviratne, S. I., van den Hurk, B., Lorenz, R., Pitman, A., Hagemann, S., Meier, A., Cheruy, F., Ducharne, A., Malyshev, S., Milly, P. C. D. (2016). Land–atmosphere feedbacks amplify aridity increase over land under global warming. *Nature Climate Change*, *6*(9), 869–874. https://doi.org/10.1038/nclimate3029

Bertolde, F., Almeida, A.-A., Pirovani, C. P., Gomes, F. P., Ahnert, D., Baligar, V. C., & Valle, R. R. (2012). Physiological and biochemical responses of Theobroma cacao L. genotypes to flooding. *Photosynthetica*, *50*(3), 447–457. https://doi.org/10.1007/s11099-012-0052-4

Bertolde, F., De Almeida, A.-A.-F., Corrêa, R. X., Gomes, F. P., Gaiotto, F. A., Baligar, V. C., & Loguercio, L. L. (2009). Molecular, physiological and morphological analysis of waterlogging tolerance in clonal genotypes of Theobroma cacao L. *Tree Physiology*, *30*(1), 56–67. https://doi.org/10.1093/treephys/tpp101

Blonder, B., Lamanna, C., Violle, C., & Enquist, B. J. (2014). The n-dimensional hypervolume. *Global Ecology and Biogeography*, *23*(5), 595–609. https://doi.org/10.1111/geb.12146

Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H., & Bollmann, K. (2013). Selecting from correlated climate variables: A major source of uncertainty for predicting species distributions under climate change. *Ecography*, *36*(9), 971–983. https://doi.org/10.1111/j.1600-0587.2013.00138.x

Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., Thuiller, W., Fortin, M.-J., Randin, C., Zimmermann, N. E., Graham, C. H., & Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, *21*(4), 481–497. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Cai, W., Borlace, S., Lengaigne, M., van Rensch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., McPhaden, M. J., Wu, L., England, M. H., Wang, G., Guilyardi, E., & Jin, F.-F. (2014). Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature Climate Change*, *4*(2), 111. https://doi.org/10.1038/nclimate2100

Clement, C. R., Levis, C., Franco-Moraes, J., & Junqueira, A. B. (2020). Domesticated nature: The culturally constructed niche of humanity. In C. Baldauf (Ed.), *Participatory biodiversity conservation* (pp. 35–51). Springer.

d'Eeckenbrugge, G. C. & Lacape, J.-M. (2014). Distribution and differentiation of wild, feral, and cultivated populations of perennial upland cotton (Gossypium hirsutum L.) in Mesoamerica and the Caribbean. *PLoS One*, *9*(9), e107458. https://doi.org/10.1371/journal.pone.0107458

Daymond, A. J. & Hadley, P. (2004). The effects of temperature and light integral on early vegetative growth and chlorophyll fluorescence of four contrasting genotypes of cacao (Theobroma cacao). *Annals of Applied Biology*, *145*(3), 257–262. https://doi.org/10.1111/j.1744-7348.2004.tb00381.x

Daymond, A. J. & Hadley, P. (2008). Differential effects of temperature on fruit development and bean quality of contrasting genotypes of cacao (Theobroma cacao). *Annals of Applied Biology*, *153*(2), 175–185. https://doi.org/10.1111/j.1744-7348.2008.00246.x

De Almeida, A. & Valle, R. R. (2007). Ecophysiology of the cacao tree. *Brazilian Journal of Plant Physiology*, *19*(4), 425–448. https://doi.org/10.1590/S1677-04202007000400011

De Almeida, J., Tezara, W., & Herrera, A. (2016). Physiological responses to drought and experimental water deficit and waterlogging of four clones of cacao (Theobroma cacao L.) selected for cultivation in Venezuela. *Agricultural Water Management*, *171*, 80–88. https://doi.org/10.1016/j.agwat.2016.03.012

de Sousa, K., van Zonneveld, M., Holmgren, M., Kindt, R., & Ordoñez, J. C. (2019). The future of coffee and cocoa agroforestry in a warmer Mesoamerica. *Scientific Reports*, *9*(1), 8828. https://doi.org/10.1038/s41598-019-45491-7

Dempewolf, H., Eastwood, R. J., Guarino, L., Khoury, C. K., Müller, J. V., & Toll, J. (2014). Adapting agriculture to climate change: A global initiative to collect, conserve, and use crop wild relatives. *Agroecology and Sustainable Food Systems*, *38*(4), 369–377. https://doi.org/10.1080/21683565.2013.870629

Derville, S., Torres, L. G., Iovan, C., & Garrigue, C. (2018). Finding the right fit: Comparative cetacean distribution models using multiple data sources and statistical approaches. *Diversity and Distributions*, *24*(11), 1657–1673. https://doi.org/10.1111/ddi.12782

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

dos Santos, I. C., de Almeida, A.-A.-F., Anhert, D., da Conceição, A. S., Pirovani, C. P., Pires, J. L., Valle, R. R., & Baligar, V. C. (2014). Molecular, physiological and biochemical responses of Theobroma

cacao L. genotypes to soil water deficit. *PLoS One*, *9*(12), e115746. https://doi.org/10.1371/journal.pone.0115746

Eitzinger, A., Farrell, A., Rhiney, K., Carmona, S., van Loosen, I., & Taylor, M. (2015). Trinidad and Tobago: Assessing the impact of climate change on cocoa and tomato. In CIAT Policy Brief No. 27. Centro Internacional de Agricultura Tropical (CIAT).

Eshetae, M. A., Hailu, B. T., & Demissew, S. (2019). Spatial characterization and distribution modelling of Ensete ventricosum (wild and cultivated) in Ethiopia. *Geocarto International*, *36*(1), 60–75. https://doi.org/10.1080/10106049.2019.1588392

Farrell, A. D., Rhiney, K., Eitzinger, A., & Umaharan, P. (2018). Climate adaptation in a minor crop species: Is the cocoa breeding network prepared for climate change? *Agroecology and Sustainable Food Systems*, *42*(7), 812–833. https://doi.org/10.1080/21683565.2018.1448924

Fremout, T., Thomas, E., Bocanegra-gonz, K. T., Aguirre-morales, C. A., Morillo-paz, A. T., Atkinson, R., Kettle, C., González-M., R., Alcázar-Caicedo, C., González, M. A., Gil-Tobón, C., Gutiérrez, J. P., Gonzalo Moscoso-Higuita, L., López-Lavalle, L. A. B., de Carvalho, D., & Muys, B. (2021). Dynamic seed zones to guide climate-smart seed sourcing for tropical dry forest restoration in Colombia. *Forest Ecology and Management*, *490*(119127), 1–15. https://doi.org/10.1016/j.foreco.2021.119127

Fremout, T., Thomas, E., Gaisberger, H., Van Meerbeek, K., Muenchow, J., Briers, S., Gutierrez-Miranda, C. E., Marcelo-Peña, J. L., Kindt, R., Atkinson, R., Cabrera, O., Espinosa, C. I., Aguirre-Mendoza, Z., & Muys, B. (2020). Mapping tree species vulnerability to multiple threats as a guide to restoration and conservation of tropical dry forests. *Global Change Biology*, *26*(6), 3552–3568. https://doi.org/10.1111/gcb.15028

Fu, Q. & Feng, S. (2014). Responses of terrestrial aridity to global warming. *Journal of Geophysical Research: Atmospheres*, *119*(13), 7863–7875. https://doi.org/10.1002/2014JD021608

Galluzzi, G., Dufour, D., Thomas, E., Van Zonneveld, M., Escobar Salamanca, A. F., Giraldo Toro, A., Rivera, A., Salazar Duque, H., Suárez Baron, H., Gallego, G., Scheldeman, X., & Gonzalez Mejia, A. (2015). An integrated hypothesis on the domestication of bactris gasipaes. *Plos One*, *10*(12), e0144644. https://doi.org/10.1371/journal.pone.0144644

Gateau-Rey, L., Tanner, E. V. J., Rapidel, B., Marelli, J.-P., & Royaert, S. (2018). Climate change could threaten cocoa production: Effects of 2015–16 El Niño-related drought on cocoa agroforests in Bahia. *Brazil. Plos ONE*, *13*(7), e0200454. https://doi.org/10.1371/journal.pone.0200454

Hällfors, M. H., Liao, J., Dzurisin, J., Grundel, R., Hyvärinen, M., Towle, K., Wu, G. C., & Hellmann, J. J. (2016). Addressing potential local adaptation in species distribution models: Implications for conservation under climate change. *Ecological Applications*, *26*(4), 1154–1169. https://doi.org/10.1890/15-0926

Hijmans, R. J. (2012). Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null model. *Ecology*, *93*(3), 679–688. https://doi.org/10.1890/11-0826.1

Hornik, K. & Böhm, W. (2020). Package clue. Version 0.3-58. https://cran.r-project.org/web/packages/clue/clue.pdf

Kindt, R. (2018). Ensemble species distribution modelling with transformed suitability values. *Environmental Modelling & Software*, *100*, 136–145. https://doi.org/10.1016/j.envsoft.2017.11.009

Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A. K., Augeri, D. M.. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*, *19*(11), 1366–1379. https://doi.org/10.1111/ddi.12096

Läderach, P., Martinez-Valle, A., Schroth, G., & Castro, N. (2013). Predicting the future climatic suitability for cocoa farming of the world's leading producer countries. *Ghana and Côte D'ivoire. Climatic Change*, *119*(3–4), 841–854. https://doi.org/10.1007/s10584-013-0774-8

Lahive, F., Hadley, P., & Daymond, A. J. (2019). The physiological responses of cacao to the environment and the implications for climate change resilience. A review. *Agronomy for Sustainable Development*, *39*(1), 5. https://doi.org/10.1007/s13593-018-0552-0

Levis, C., Costa, F. R. C., Bongers, F., Peña-Claros, M., Clement, C. R., Junqueira, A. B., & ter Steege, H. (2017). Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science*, *355*(6328), 925–931. https://doi.org/10.1126/science.aal0157

Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, *28*(3), 385–393. https://doi.org/10.1111/j.0906-7590.2005.03957.x

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2019). *Cluster: Cluster analysis basics and extensions. R package version 2.0. 8.* : R Foundation for Statistical Computing.

Mateo, R. G., Croat, T. B., Felicísimo, Á. M., & Munoz, J. (2010). Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Diversity and Distributions*, *16*(1), 84–94. https://doi.org/10.1111/j.1472-4642.2009.00617.x

Medina, V. & Laliberte, B. (2017). A review of research on the effects of drought and temperature stress and increased CO2 on Theobroma cacao L., and the role of genetic diversity to address climate change. Bioversity International.

Miller, A. J. & Knouft, J. H. (2006). GIS-based characterization of the geographic distributions of wild and cultivated populations of the Mesoamerican fruit tree Spondias purpurea (Anacardiaceae). *American Journal of Botany*, *93*(12), 1757–1767. https://doi.org/10.3732/ajb.93.12.1757

Miller, R. P. & Nair, P. K. R. (2006). Indigenous agroforestry systems in Amazonia: From prehistory to today. *Agroforestry Systems*, *66*(2), 151–164. https://doi.org/10.1007/s10457-005-6074-1

MINAGRI. (2018). Análisis de la cadena productiva del cacao con enfoque en los pequeños productores de limitado acceso al mercado. Ministerio de Agricultura y Riego.

Morton, J. F. (2007). The impact of climate change on smallholder and subsistence agriculture. *Proceedings of the National Academy of Sciences*, *104*(50), 19680–19685. https://doi.org/10.1073/pnas.0701855104

Moser, G., Leuschner, C., Hertel, D., Hölscher, D., Köhler, M., Leitner, D., Michalzik, B., Prihastanti, E., Tjitrosemito, S., & Schwendenmann, L. (2010). Response of cocoa trees (Theobroma cacao) to a 13-month desiccation period in Sulawesi. *Indonesia. Agroforestry Systems*, *79*(2), 171–187. https://doi.org/10.1007/s10457-010-9303-1

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

Ofori, A., Konlan, S., Dadzie, M. A., & Amoah, F. M. (2014). Genotypic performance of cocoa (Theobroma cacao L.) during establishment under natural drought stress. *Journal of Crop Improvement*, *28*(6), 804–824. https://doi.org/10.1080/15427528.2014.947529

Parra-Quijano, M., Iriondo, J. M., & Torres, E. (2012). Ecogeographical land characterization maps as a tool for assessing plant adaptation and their implications in agrobiodiversity studies. *Genetic Resources and Crop Evolution*, *59*(2), 205–217. https://doi.org/10.1007/s10722-011-9676-7

Parry, M. L., Rosenzweig, C., Iglesias, A., Livermore, M., & Fischer, G. (2004). Effects of climate change on global food production under SRES emissions and socio-economic scenarios. *Global Environmental Change*, *14*(1), 53–67. https://doi.org/10.1016/j.gloenvcha.2003.10.008

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence

data. *Ecological Applications*, *19*(1), 181–197. https://doi. org/10.1890/07-2153.1

Porter, J. R., Xie, L., Challinor, A. J., Cochrane, K., Howden, S. M., Iqbal, M. M., Lobell, D. B. & Travasso, M. I. (2014). Food security and food production systems. In *Climate change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. Contribution of working group ii to the fifth assessment report of the intergovernmental panel on climate change* (pp. 485–533). Cambridge University Press.

Ramirez Villegas, J. & Jarvis, A. (2010). Downscaling global circulation model outputs: The delta method. Decision and policy analysis working paper No. 1. International Center for Tropical Agriculture (CIAT).

Ramirez-Villegas, J., Jarvis, A., & Läderach, P. (2013). Empirical approaches for assessing impacts of climate change on agriculture: The EcoCrop model and a case study with grain sorghum. *Agricultural and Forest Meteorology*, *170*, 67–78. https://doi.org/10.1016/j.agrfo rmet.2011.09.005

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. *Science of the Total Environment*, *556*, 231–241. https://doi.org/10.1016/j.scito tenv.2016.03.024

Segurado, P. & Araujo, M. B. (2004). An evaluation of methods for modelling species distributions. *Journal of Biogeography*, *31*(10), 1555–1568. https://doi.org/10.1111/j.1365-2699.2004.01076.x

Sekula, M., Datta, S., & Datta, S. (2020). optCluster: determine optimal clustering algorithm and number of clusters. Version 1.3.0. R Package Version, *1*(1). https://cran.r-project.org/web/packages/optCluster/ optCluster.pdf

Smit, B. & Skinner, M. W. (2002). Adaptation options in agriculture to climate change: A typology. *Mitigation and Adaptation Strategies for Global Change*, *7*(1), 85–114. https://doi.org/10.1023/A:10158 62228270

Smith, A. B., Alsdurf, J., Knapp, M., Baer, S. G., & Johnson, L. C. (2017). Phenotypic distribution models corroborate species distribution models: A shift in the role and prevalence of a dominant prairie grass in response to climate change. *Global Change Biology*, *23*(10), 4365–4375. https://doi.org/10.1111/gcb.13666

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., & Midgley, P. M. (2013). *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Thomas, E. (2017). Do multiple origins of domestication matter for Amazon Forest? Response to Levis et al Persistent effects of pre-Columbian plant domestication on Amazonian forest composition, Science 03 Mar 2017. *Science ELetter*, *355*(6328), 925–931. http:// science.sciencemag.org/content/355/6328/

Thomas, E., Alcázar Caicedo, C., Loo, J., & Kindt, R. (2014). The distribution of the Brazil nut (Bertholletia excelsa) through time: From range contraction in glacial refugia, over human-mediated expansion, to anthropogenic climate change. *Boletim do Museu Paraense Emílio Goeldi Ciências Naturais*, *9*, 267–291.

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

Thuiller, W. (2003). BIOMOD–optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, *9*(10), 1353–1362. https://doi. org/10.1046/j.1365-2486.2003.00666.x

Thuiller, W., Guéguen, M., Renaud, J., Karger, D. N., & Zimmermann, N. E. (2019). Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications*, *10*(1), 1–9. https://doi.org/10.1038/s41467-019-09519-w

Tribett, W. R., Salawitch, R. J., Hope, A. P., Canty, T. P., & Bennett, B. F. (2017). Paris INDCs (Ch. 3). In R. J. Salawitch, T. P. Canty, A. P. Hope, W. R. Tribett, & B. F. Bennett (Eds.), *Paris climate agreement: Beacon of hope* (pp. 115–146). Springer International Publishing.

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). block CV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, *10*(2), 225–232. https:// doi.org/10.1111/2041-210X.13107

Wenger, S. J. & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, *3*(2), 260–267. https://doi. org/10.1111/j.2041-210X.2011.00170.x

Wood, G. A. R. & Lass, R. A. (2008). *Cocoa*. John Wiley & Sons.

Wunderlich, R. F., Lin, Y.-P., Anthony, J., & Petway, J. R. (2019). Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. *Nature Conservation*, *35*, 97–117. https://doi.org/10.3897/natureconservation.35.33918

Zuidema, P. A., Leffelaar, P. A., Gerritsma, W., Mommer, L., & Anten, N. P. R. (2005). A physiological production model for cocoa (Theobroma cacao): Model presentation, validation and application. *Agricultural Systems*, *84*(2), 195–225. https://doi.org/10.1016/j. agsy.2004.06.015

BIOSKETCH

Viviana Ceccarelli is a research fellow at the Alliance of Bioversity International—CIAT, interested in species distribution modelling and other spatial analysis to support adaptation to climate change and forest restoration and conservation.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.