



OPEN

# Global distribution of soapberries (*Sapindus* L.) habitats under current and future climate scenarios

Jiming Liu<sup>1,2,7</sup>, Lianchun Wang<sup>3,7</sup>, Caowen Sun<sup>4</sup>, Benye Xi<sup>1</sup>, Doudou Li<sup>1</sup>, Zhong Chen<sup>1,2</sup>, Qiuyang He<sup>5</sup>, Xuehuang Weng<sup>6</sup> & Liming Jia<sup>1,2</sup>✉

*Sapindus* (*Sapindus* L.) is a widely distributed economically important tree genus that provides biodiesel, biomedical and biochemical products. However, with climate change, deforestation, and economic development, the diversity of *Sapindus* germplasms may face the risk of destruction. Therefore, utilising historical environmental data and future climate projections from the BCC-CSM2-MR global climate database, we simulated the current and future global distributions of suitable habitats for *Sapindus* using a Maximum Entropy (MaxEnt) model. The estimated ecological thresholds for critical environmental factors were: a minimum temperature of 0–20 °C in the coldest month, soil moisture levels of 40–140 mm, a mean temperature of 2–25 °C in the driest quarter, a mean temperature of 19–28 °C in the wettest quarter, and a soil pH of 5.6–7.6. The total suitable habitat area was  $6059.97 \times 10^4$  km<sup>2</sup>, which was unevenly distributed across six continents. As greenhouse gas emissions increased over time, the area of suitable habitats contracted in lower latitudes and expanded in higher latitudes. Consequently, surveys and conservation should be prioritised in southern hemisphere areas which are in danger of becoming unsuitable. In contrast, other areas in northern and central America, China, and India can be used for conservation and large-scale cultivation in the future.

*Sapindus* (*Sapindus* L.) is a genus containing 13 evergreen and deciduous tree species in the Sapindaceae family. It is globally distributed across warm-temperate and tropical regions in Southeast Asia and North and South America<sup>1</sup>. *S. saponaria* L. is the type species of the *Sapindus* genus, and it is widely distributed in North and South America, with some localised distributions in Africa and Australia. *S. mukorossi* Gaertn. is the second most widespread species and is found in East and Southeast Asia. *Sapindus* seed oil is high yielding (26.15–44.69%) with a high medium-chain monounsaturated fatty acid content<sup>2</sup>. Crude extracts from the fruit pericarps of *Sapindus* are also rich in triterpenoid saponins (4.14–27.04%) and sesquiterpenoids<sup>3</sup>, such as Saponin A, mukurozi-saponin G, and saponin K<sup>4</sup>, which exhibit excellent surface activity as well as antibacterial<sup>5</sup>, elution<sup>6–8</sup>, antibacterial<sup>9,10</sup>, insecticidal<sup>11</sup>, pharmacological<sup>12</sup>, and physiological<sup>13</sup> effects. Saponin serves as an efficient natural surfactant in commercial soaps, shampoos, and cosmetic cleansers<sup>14</sup>. In recent years, *Sapindus* soap products have been sold in the US, Europe and China, with a market of over 10 million sales in China. Therefore, *Sapindus* species are regarded as economically important sources of biodiesel, as well as biomedical and multi-functional products<sup>1,15</sup>. *Sapindus* germplasm resources are generally scattered in the form of single plants or extremely small populations. With global deforestation and rapid economic development, the diversity of *Sapindus* have been persistently damaged or lost<sup>1</sup>. Modern cultivation of *Sapindus* species has only begun recently and lacks support from relevant research. It is still in a state of low yield with elite varieties lacking and severe germplasm destruction ongoing<sup>1,4</sup>. We suspect that the diversity and habitat of *Sapindus* will suffer further damage in the background of future climate change. Therefore, the protection and management of the core *Sapindus* distribution areas should be strengthened, and natural populations at risk of destruction should be protected through in situ or ex situ conservation efforts.

<sup>1</sup>Key Laboratory of Silviculture and Conservation of the Ministry of Education, Beijing Forestry University, 35 E Qinghua Rd., Beijing 100083, People's Republic of China. <sup>2</sup>National Energy R&D Center for Non-Food Biomass, Beijing Forestry University, Beijing 100083, People's Republic of China. <sup>3</sup>College of Forestry, Southwest Forestry University, Kunming Yunnan 650224, People's Republic of China. <sup>4</sup>College of Forestry, Nanjing Forestry University, Nanjing, Jiangsu 210037, People's Republic of China. <sup>5</sup>China Jiliang University, Hangzhou, Zhejiang 310018, People's Republic of China. <sup>6</sup>Yuanhua Forestry Biological Technology Co., Ltd., Nanming, Fujian 650216, People's Republic of China. <sup>7</sup>These authors contributed equally: Jiming Liu and Lianchun Wang. ✉email: jlm@bjfu.edu.cn

Code	Environmental factor	Percent contribution (%)	Suitable threshold	Units
Bio6	Min temperature of coldest month	26	0–20	°C
Sm	Soil moisture	14.1	40–140	mm
Bio9	Mean temperature of driest quarter	13.1	2–25	°C
Bio8	Mean temperature of wettest quarter	11.6	19–28	°C
Sph	Soil pH	10.5	5.6–7.6	

**Table 1.** Contributions and thresholds of the major environmental factors in the MaxEnt models for the current suitable habitat of *Sapindus*.

Niches are habitats with the minimum thresholds necessary for survival<sup>16</sup>. The forest niche is strongly affected by the environment, and it changes or moves with environmental change. The Intergovernmental Panel on Climate Change (IPCC) estimates that a 0.2 °C temperature increase will occur in each future decade that is subject to greenhouse gas emissions<sup>17</sup>. Temperatures will rise by a maximum of 2.6–4.8 °C or a minimum of 0.3–1.7 °C in the twenty-first century<sup>17</sup>. In the face of upcoming rapid climate change, forest trees will unlikely be able to adjust their range with sufficient speed to colonise suitable areas. Species extinction rates would subsequently increase, and warmer temperatures may impact plant growth and yield<sup>18–20</sup>. Hence, there is an urgent need to understand the extent of climate change in the coming decades, and the use of alternative methods to assess its impact on forest tree habitats will be helpful for designing conservation and cultivation plans in the future<sup>21,22</sup>.

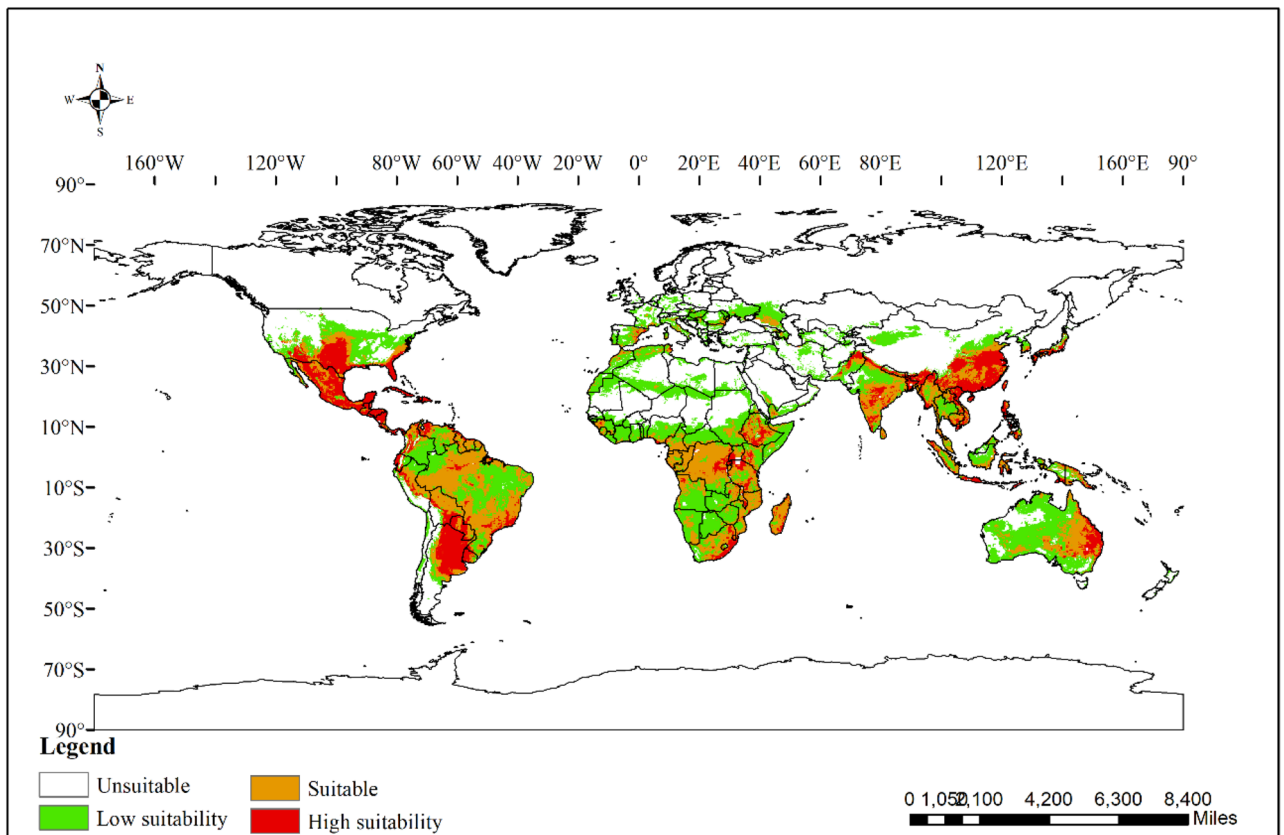
Species distribution modelling is an emerging research field based on niche theory. Its principle is to infer the ecological needs of species through mathematical models based on occurrence data and environmental variables, and to create a statistical or mechanistic model of its potential distribution<sup>22–24</sup>. At present, the commonly used niche models are GARP (Genetic Algorithm for Rule-set Production)<sup>25</sup>, MaxEnt (Maximum entropy modeling)<sup>26</sup>, Bioclim<sup>27</sup>, Random Forest<sup>28</sup>, and the Boosted Regression Tree<sup>29</sup>. Many model intercomparison studies have reported that the MaxEnt model, which is based on the principle of maximum entropy<sup>23,26,30</sup>, typically outperforms other species distribution models (SDMs) in terms of high tolerance and high predictive accuracy<sup>31–33</sup>. Over the past 10 years, worldwide research teams have achieved excellent results in the study of rare animal and plant diversity protection<sup>34–38</sup>, invasive species risk prediction<sup>39–41</sup>, marine ecosystem protection<sup>42,43</sup>, disaster distribution prediction<sup>44</sup>, and disease propagation<sup>45,46</sup> using the MaxEnt model. Zhang et al.<sup>47</sup> found that the area of suitable range of *Cinnamomum camphora* (L.) Presl will increase and continue to move to the northwest of China under future climate change scenarios. Peng et al.<sup>48</sup> simulated the distribution of suitable areas for the new oil crop *Paeonia ostii* and found that *Paeonia ostii* will increase its suitable area at high latitudes while decrease at low latitudes. And there were a decreasing tendency observed for three *Coptis* herbal species under the current and future habitat distributions<sup>49</sup>. It is noteworthy that most previous studies have focused on the prediction of suitable habitats in local regions with models that basically follow the default parameters, resulting in suboptimal model accuracy<sup>48,50</sup>. At present, there are no reports or studies, to our knowledge, that focus on current and future suitable habitat projections on a global scale for *Sapindus*.

Based on 5674 global occurrence data records for *Sapindus*, historical environmental factors, and future climate models combined with ArcGIS 10.5 and MaxEnt modelling, we predicted the current global distribution of suitable habitats for *Sapindus* and its response to future climate change scenarios. This study had three objectives: (1) to evaluate the main environmental factors affecting the distribution of *Sapindus*, (2) to explore the distribution of suitable habitats for *Sapindus* under current environmental conditions, and (3) to predict the redistribution pattern of potential *Sapindus* habitats in response to future climate change scenarios in the twenty-first century, and (4) to identify hotspots of habitat degradation and expansion to facilitate climate change-adaptive biological conservation recommendations.

## Results

**Modelling evaluation and selection.** We implemented model performance screening using the SDM-toolbox data analysis package in ArcGIS. After simulating regularisation multipliers of 0.5, 1, and 2, combined with five feature type (linear, product, hinge, quadratic, and threshold) model parameters, model performance was assessed based on the AUC values. According to the evaluation results, we found that the MaxEnt model with a regularisation multiplier of 1 and a feature type of LQ (linear and quadratic) parameters was the most successful. Therefore, we applied this parameter model to conduct further analysis. The MaxEnt model for current *Sapindus* suitable habitats provided satisfactory results, with an TSS and AUC value of 0.811 and 0.835, respectively. And The MaxEnt model for future *Sapindus* suitable habitats also achieved satisfactory results, with AUC values ranging from 0.838 to 0.848, and TSS values ranging from 0.810 to 0.821.

**Current suitable habitats for *Sapindus*.** The MaxEnt model's internal jackknife test of factor importance showed that the minimum temperature of the coldest month (Bio6, 26.0% of variation), soil moisture (Sm, 14.1% of variation), mean temperature of the driest quarter (Bio9, 13.1% of variation), mean temperature of the wettest quarter (bio8, 11.6% of variation), and soil pH (Sph, 10.5% of variation) were the major contributors to the *Sapindus* distribution model, with a cumulative contribution of 75.3% (Table 1). Soil carbon, annual mean temperature (Bio 1), annual mean UVb (AmUV) and 13 other indicators contributed 24.7% of the cumulative contribution. According to the MaxEnt results and environmental factor response curves (Fig. S1), the eco-



**Figure 1.** Map of suitable habitat areas for *Sapindus* under current climate conditions. Created in ESRI ArcMap 10.5 (<https://support.esri.com/en/Products/Desktop/arcgis-desktop/arcmap>).

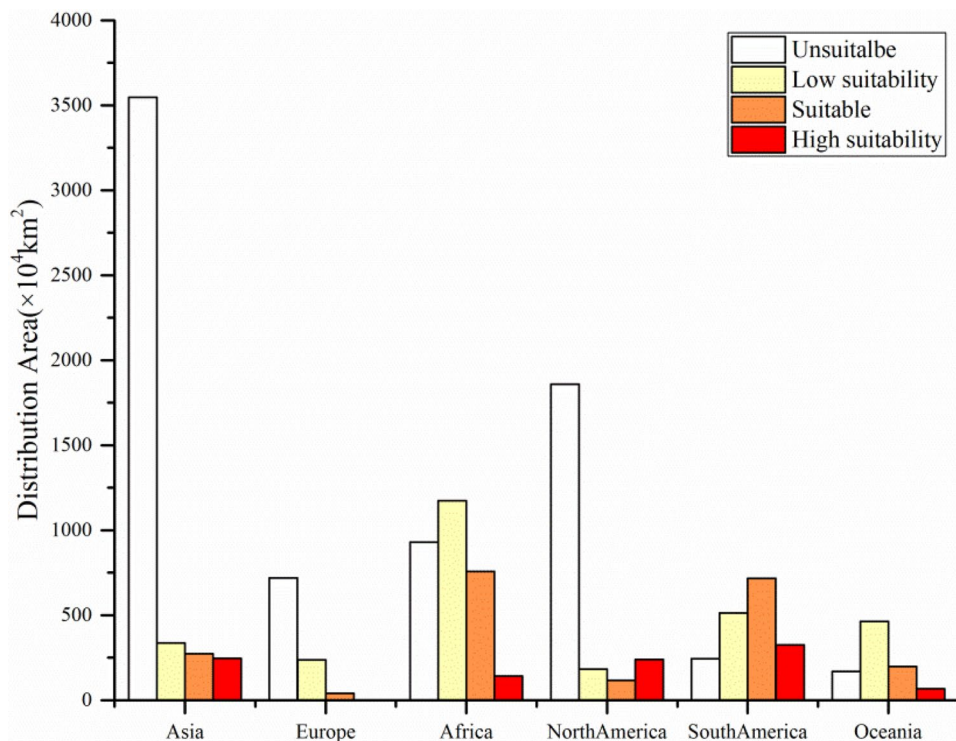
Suitable level	Distribution area ( $10^4$ km <sup>2</sup> )	Distribution area ratio (%)
Unsuitable	8840.03	59.33
Low suitability	2843.00	19.08
Suitable	2092.92	14.05
High suitability	1124.04	7.54

**Table 2.** Area and proportion of different suitability levels for global *Sapindus* habitats under the current environment.

logical thresholds for the critical environmental factors were the minimum temperature of the coldest month (0–20 °C), soil moisture (40–140 mm), mean temperature of the driest quarter (2–25 °C), mean temperature of the wettest quarter (19–28 °C), and soil pH (5.6–7.6).

The suitable habitats for *Sapindus* (Fig. 1) were widely distributed across six continents, except Antarctica, and were mainly distributed in southern and south-eastern Asia, southern North America, northern and central South America, central and southern Africa, and eastern Oceania. The area of total suitable habitat was  $6059.97 \times 10^4$  km<sup>2</sup>, accounting for 40.67% of the global land area. Among them, low suitability areas accounted for a relatively high area of  $2843.00 \times 10^4$  km<sup>2</sup>, accounting for 19.08% of the global land area. The suitable and high suitability areas were  $2092.92 \times 10^4$  km<sup>2</sup> (14.05%) and  $1124.04 \times 10^4$  km<sup>2</sup> (7.54%), respectively (Table 2).

Suitable habitats for *Sapindus* were unevenly distributed across six continents (Fig. 2). The highest area of total suitable habitat was in Africa ( $2072.258 \times 10^4$  km<sup>2</sup>), which contained mainly low suitability areas ( $1174.139 \times 10^4$  km<sup>2</sup>) and a small amount of suitable areas ( $756.24 \times 10^4$  km<sup>2</sup>) and high suitability ( $141.88 \times 10^4$  km<sup>2</sup>) areas. The second highest area was in South America, with a total suitable habitat area of  $1,557.337 \times 10^4$  km<sup>2</sup> composed of suitable areas ( $717.43 \times 10^4$  km<sup>2</sup>) and high suitability ( $326.21 \times 10^4$  km<sup>2</sup>) areas. The suitable habitat area for *Sapindus* in Asia was  $853.61 \times 10^4$  km<sup>2</sup>, of which the suitable and high suitability areas were  $272.20 \times 10^4$  km<sup>2</sup> and  $245.13 \times 10^4$  km<sup>2</sup>, respectively. The suitable habitat area for *Sapindus* in Europe was the lowest of all at only  $281.24 \times 10^4$  km<sup>2</sup>, of which the suitable and high suitability areas were  $41.20 \times 10^4$  km<sup>2</sup> and  $2.05 \times 10^4$  km<sup>2</sup>, respectively. The high suitability habitats for *Sapindus* were concentrated in south-central South America, including Argentina, Paraguay, and Uruguay; southern North America, including the southern United States, Mexico, and Central American countries; south-eastern Asia, including southern China, northern Vietnam, Bangladesh,



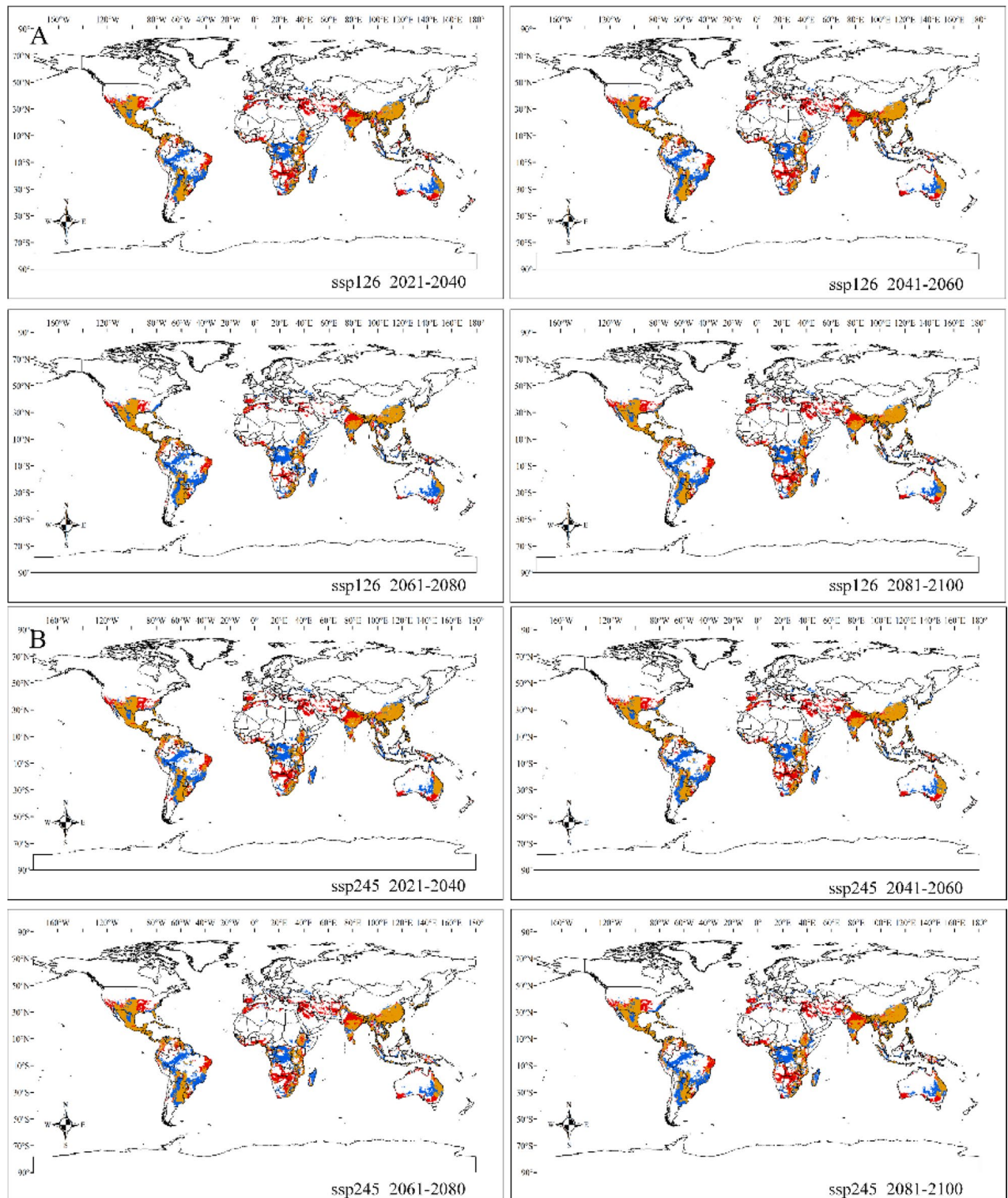
**Figure 2.** Distribution area of *Sapindus* habitats on six continents under the current environment.

northern and central India, Nepal, and northern Pakistan; and very few areas of eastern Oceania and central Africa.

**Potential distribution of *Sapindus* under future climate conditions.** By comparing the current suitable habitats (Fig. 1) with the projected suitable habitats from 2020 to 2100, we predicted the potential redistribution of *Sapindus* habitats in response to climate change in the twenty-first century under four climate scenarios (Fig. 3). Different trends emerged under the different future climate scenarios. The expansion area of total suitable habitat ranged from  $607.45 \times 10^4 \text{ km}^2$  (ssp370, 2061–2080) to  $1092.86 \times 10^4 \text{ km}^2$  (ssp585, 2081–2100), and the contraction area ranged from  $1041.73 \times 10^4 \text{ km}^2$  (ssp245, 2021–2040) to  $1267.23 \times 10^4 \text{ km}^2$  (ssp370, 2081–2100). Overall, there was a significant contraction in the size of suitable habitats with increasing greenhouse gas emissions. It is noteworthy that the contraction was most pronounced in the second half of the twenty-first century in the southern hemisphere, including South America, Central Africa, and Oceania.

There was significant expansion and contraction of suitable habitats for *Sapindus* in all the future climate scenarios and these expansions and contractions differed significantly between continents (Table S1). Intriguingly, suitable habitats in Asia, Europe, and North America all showed expansion at higher latitudes, and the expansion of Asian habitats ranged from  $164.84 \times 10^4 \text{ km}^2$  (ssp126, 2061–2080) to  $293.40 \times 10^4 \text{ km}^2$  (ssp245, 2061–2080). European habitat expansion ranged from  $3.94 \times 10^4 \text{ km}^2$  (ssp370, 2061–2080) to  $32.34 \times 10^4 \text{ km}^2$  (ssp245, 2020–2040), while expansion of the North American habitats ranged from  $79.19 \times 10^4 \text{ km}^2$  (ssp370, 2061–2080) to  $127.39 \times 10^4 \text{ km}^2$  (ssp245, 2020–2040). The expansion of Asian habitats mainly occurred in northern India, Afghanistan, and the Middle East. The European habitat expansion areas were mainly located in Spain, Portugal, Italy, and Greece, and the North American habitat expansion areas spanned from the southern United States to the central part of the country, including California, Arkansas, Tennessee, and Missouri (Fig. 3).

Conversely, suitable habitats for *Sapindus* in the southern hemisphere contracted significantly, with South American habitat areas shrinking by  $405.54 \times 10^4 \text{ km}^2$  (ssp126, 2041–2060) to  $471.38 \times 10^4 \text{ km}^2$  (ssp370, 2081–2100), which was 2.96 to 3.97 times the expansion area ( $118.79 \times 10^4 \text{ km}^2$  [ssp126, 2061–2080] to  $136.80 \times 10^4 \text{ km}^2$  [ssp245, 2020–2040]). Oceania contracted by  $94.46 \times 10^4 \text{ km}^2$  (ssp245, 2020–2040) to  $159.70 \times 10^4 \text{ km}^2$  (ssp370, 2081–2100), which was 0.74 to 15.63 times the expansion area ( $10.22 \times 10^4 \text{ km}^2$  [ssp370, 2061–2080] to  $126.73 \times 10^4 \text{ km}^2$  [ssp585, 2081–2100]). African habitat areas contracted by  $324.27 \times 10^4 \text{ km}^2$  (ssp585, 2041–2060) to  $402.50 \times 10^4 \text{ km}^2$  (ssp370, 2081–2100), which was 0.86 to 2.11 times the expansion area ( $191.19 \times 10^4$  [ssp370, 2061–2080] to  $378.75 \times 10^4 \text{ km}^2$  [ssp585, 2081–2100]) (Table S1). The contracted areas were mainly located in Brazil, Peru, Bolivia, Paraguay, and Argentina in South America; the Congo, Mozambique, and Madagascar in Africa; and eastern Australia in Oceania (Fig. 3).



**Figure 3.** Changes in the distribution of potential suitable habitats for *Sapindus* under the ssp126 (A), ssp245 (B), ssp370 (C), and ssp585 (D) scenarios from years 2020 to 2100 compared with the current distribution of suitable habitats. Note: Red areas indicate habitat expansion, yellow areas indicate no change and blue areas indicate habitat contraction. Created in ESRI ArcMap 10.5 (<https://support.esri.com/en/Products/Desktop/arcgis-desktop/arcmap>).

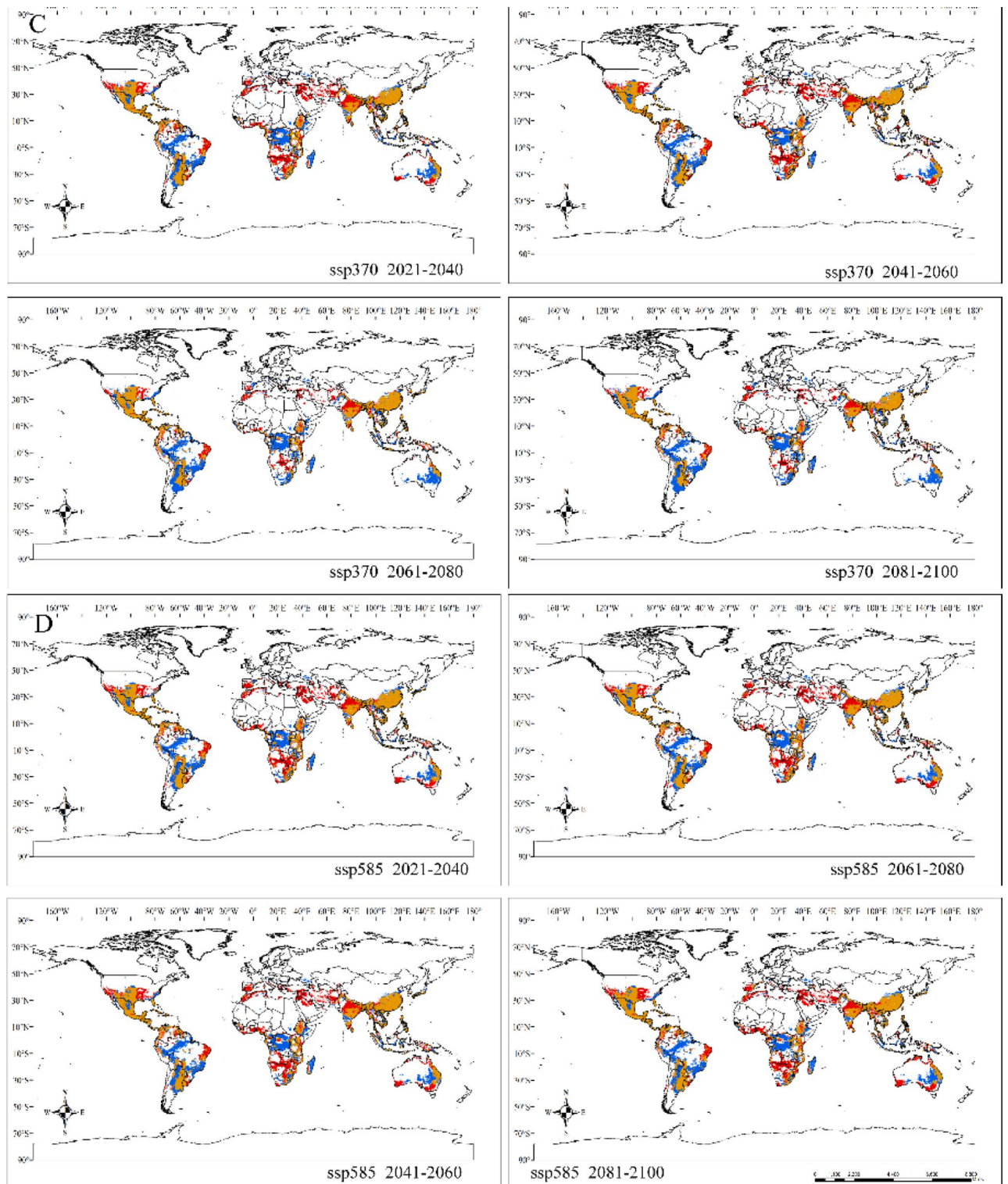


Figure 3. (continued)

### Discussion

There are 13 species of *Sapindus* worldwide, all of which are scattered in ecosystems as either individual plants or small populations<sup>1</sup>. With global deforestation, rapid expansion of industrial, agricultural, and urban land use, increasing human interference, and over-exploitation of groundwater in the last century, *Sapindus* germplasm resources have been persistently damaged or lost<sup>3</sup>. Therefore, it is imperative to sustainably exploit and conserve *Sapindus* germplasm diversity by predicting the distribution of current and future suitable habitats through ecological niche modelling.

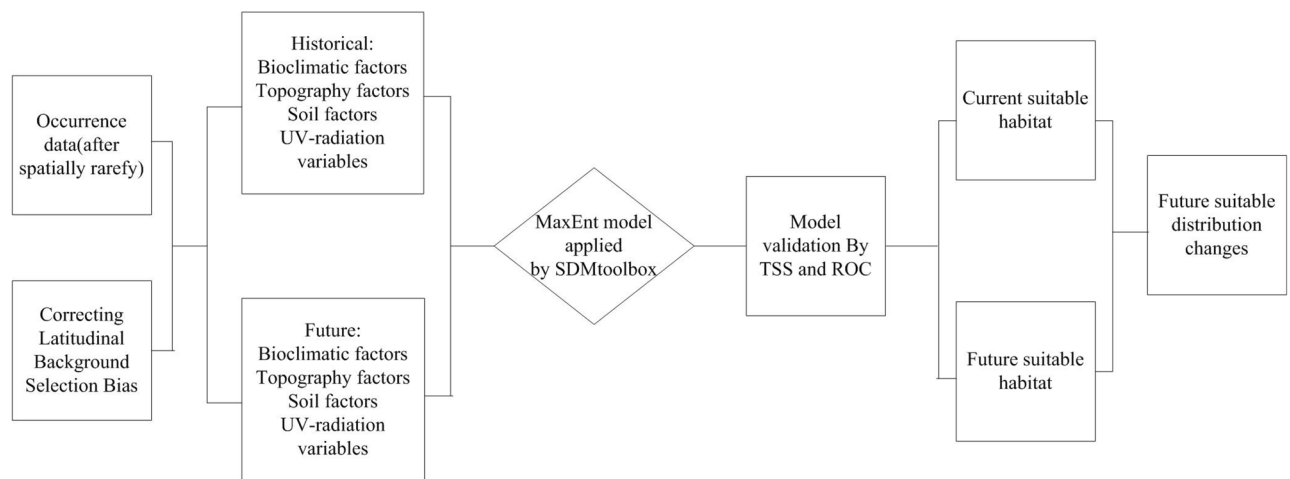
This study analysed projections of current and future distributions of suitable habitats for *Sapindus* using the MaxEnt model. Habitat maps were created for *Sapindus* based on occurrence data sets with great TSS and AUC values for the present and the future, respectively. Therefore, we consider that our model performance is robust and adequate for construing the overall suitable habitat distribution of *Sapindus*. To our knowledge, this is the first study to analyse the suitable habitat distribution of *Sapindus* for the current and future using the MaxEnt model.

**Suitable habitat distribution patterns of *Sapindus* under the current environment.** After a long period of natural selection, anthropogenic disturbance, and geographic isolation, genetic variation in the *Sapindus* genus has increased greatly, resulting in patterns of geographic variation in both phenotype and ecological adaptation<sup>3,51,52</sup>. According to the response curves of ecological factors, the relationships between the probability of species occurrence and the main ecological factors can be determined, and it is generally considered that if the probability of species occurrence is greater than 60%, the corresponding ecological factor thresholds are suitable for the survival of this species<sup>30,53,54</sup>. In our study, the MaxEnt results and environmental factor response curves indicated that the critical environmental factors affecting suitable habitats for *Sapindus* were minimum temperature of coldest month, soil moisture, mean temperature of driest quarter, mean temperature of wettest quarter, and soil pH. Similarly, Adeyemi found that the minimum temperature of the coldest month was the most critical environmental factor in determining the suitable habitat distribution of Sapindaceae Juss. species in West Africa<sup>55</sup>. Sun et al.<sup>56,57</sup> also found that the annual minimum temperature played an important role in *Sapindus* saponin variation, but high precipitation levels were found to inhibit saponin synthesis<sup>55–57</sup>. Liu found that the precipitation of the warmest quarter and isothermality were the critical environmental factors that determine the distribution of *Sapindus* habitats in China. *Sapindus* is highly thermophilous<sup>3,58,59</sup> and xerophilous tree species<sup>1</sup> that prefer areas with abundant heat and sunlight resources<sup>3</sup>. They do not tolerate excessive cold and humidity. Moderate rainfall and soil conditions are more conducive to their survival<sup>60</sup>. On a global scale, we found that the critical environmental factor for *Sapindus* survival was the minimum temperature of the coldest month. However, extremely low temperatures (below  $-10^{\circ}\text{C}$ ) are also detrimental to the survival of *Sapindus*<sup>61</sup>. This indicates that the current suitable habitats for *Sapindus* are likely to be mainly distributed in tropical and subtropical regions between the  $50^{\circ}$  north parallel and the  $40^{\circ}$  south parallel latitudes. This is consistent with the findings of Liu<sup>60</sup>, who found that the northernmost margin of *Sapindus* distribution in China did not go beyond the Qinling Mountains or the Huai River (for the most part). *Sapindus* is deep-rooted and drought-tolerant<sup>62</sup> tree species that does not require high levels of soil moisture<sup>63</sup>. We have found that soils with a soil moisture ranges of 40 to 140 mm and weak acidity or neutrality (soil Ph 5.6–7.6) are suitable for the survival of *Sapindus*.

Our results indicated that the total suitable habitat area for *Sapindus* globally was  $6059.97 \times 10^4$  km<sup>2</sup>. Of this, the proportion of low suitability areas accounted for 46.91%, suitable areas accounted for 34.54%, and high suitability areas accounted for 18.55%. The total suitable habitats for *Sapindus* were found to be unevenly distributed across six continents, with high suitability areas concentrated in the Mississippi Plain, the Florida Peninsula, and Mexico in North America; La Plata Plain in South America; and the North China Plain, Middle and Lower Yangtze Plain, Sichuan Basin, Yunnan-Guizhou Plateau, and Southern Himalayas in Asia. These results are consistent with a previous study, which found that *Sapindus* are mainly distributed in America and Asia<sup>1,64</sup>. High suitability areas are more conducive to cultivation and the conservation of *Sapindus* diversity than low suitability and suitable habitats. Therefore, it is recommended that ex situ conservation, cultivation, and breeding of *Sapindus* species be implemented in high-suitability areas.

**Response of suitable habitats for *Sapindus* to future climate change.** Climate plays a significant role in defining species distributions, and changes in the distribution of species are also the most clear and direct response to climate. Global warming may substantially change the structure and function of terrestrial ecosystems, resulting in significant changes in the extent and distribution of biological habitats<sup>16,20</sup>.

Our results indicated that the size and distribution of suitable habitats for *Sapindus* varied under different climatic scenarios, suggesting that climate change had an uncertain and region-specific effect on the distribution of these habitats. Under future climate scenarios, we estimate that suitable habitats for *Sapindus* will expand and contract significantly across continents. These habitats will significantly contract in lower latitudes and expand in higher latitudes. Our model showed that as greenhouse gas emissions increased and time passed, the area of suitable habitats contracted sharply, peaking in the second half of the twenty-first century (2081–2100). In terms of the northern and southern hemispheres, the *Sapindus* suitable habitats for *Sapindus* in the northern hemisphere predominantly expanded, while in the southern hemisphere they predominantly contracted. Jayasinghe and Prevéy also found a declining trend in the area of suitable habitats for *Camellia sinensis* (L.) O. Kuntze<sup>65</sup> and huckleberry<sup>54</sup> in the face of global warming, while He<sup>66</sup> found an expanding trend in the highly suitable habitat area for *Xanthoceras sorbifolia* Bunge under global warming in China. The *Sapindus* genus consists of drought-tolerant and hardy species, but they are more sensitive to extreme temperatures. With future global warming, parts of the high latitudes of the northern and southern hemispheres may be transformed into suitable habitats for *Sapindus* due to an increase in extreme low temperatures, while parts of the lower latitudes may experience persistent extreme heat or drought as the climate continues to warm. This poses a major problem for *Sapindus* survival. In He's study<sup>66</sup>, the distribution of *X. sorbifolia* was observed to be localised in areas of the Loess Plateau in China, which is at a higher latitude. Thus, a trend toward the expansion of suitable cultivation areas in the face of global warming is already evident. Our results indicated that *Sapindus* will also experience some localised suitable habitat expansions in the northern hemisphere at higher latitudes, but, generally, *Sapindus* habitats will predominantly contract in the face of global warming. The southern hemisphere, including the Congo Basin and Madagascar Island in Africa; the Amazon Plains, the Brazilian Plateau, and La Plata Plain in South America; and



**Figure 4.** Design framework of the suitable habitat distribution model for *Sapindus*.

the central plains of Oceania, may no longer be suitable for *Sapindus* in the face of future global warming. These areas should be prioritised for surveys and collection of *Sapindus* germplasms<sup>3,15,56</sup>. Some elite or distinctive germplasms could be ex situ-conserved by vegetative propagation in order to conserve as much *Sapindus* genetic diversity as possible. The Mississippi Plain, the Florida Peninsula, Mexico, Central America, the Middle and Lower Yangtze Plain, the Yungui Plateau, and central India, which are relatively less affected by climate change, can be used as a base for resource conservation, large-scale cultivation, and utilisation of *Sapindus* in the future.

Although the MaxEnt model is the most commonly used model in recent years for predicting species distribution change and has greater prediction performance than other species distribution prediction models, it still has some limitations<sup>53</sup>. Despite climatic, soil, topographical, and solar radiation factors being taken into account in this model, species distributions can also be constrained by other factors such as adaptive capacity, interspecific interactions, human activities, land use. When all factors are considered together, we estimate that the area of suitable habitats for *Sapindus* will further reduce and the contraction in habitat area in the face of future global warming may be more significant. However, importing all variables into the model may lead to difficulties in variable screening, the effects of key variables may be weakened, and the resulting simulations may not necessarily be more accurate than the current model. Therefore, our findings on the distribution pattern of suitable habitats for *Sapindus* and its response to future climate change, based on the MaxEnt model, provide an important theoretical basis and valuable recommendations for the conservation and sustainable exploitation of *Sapindus* genetic diversity.

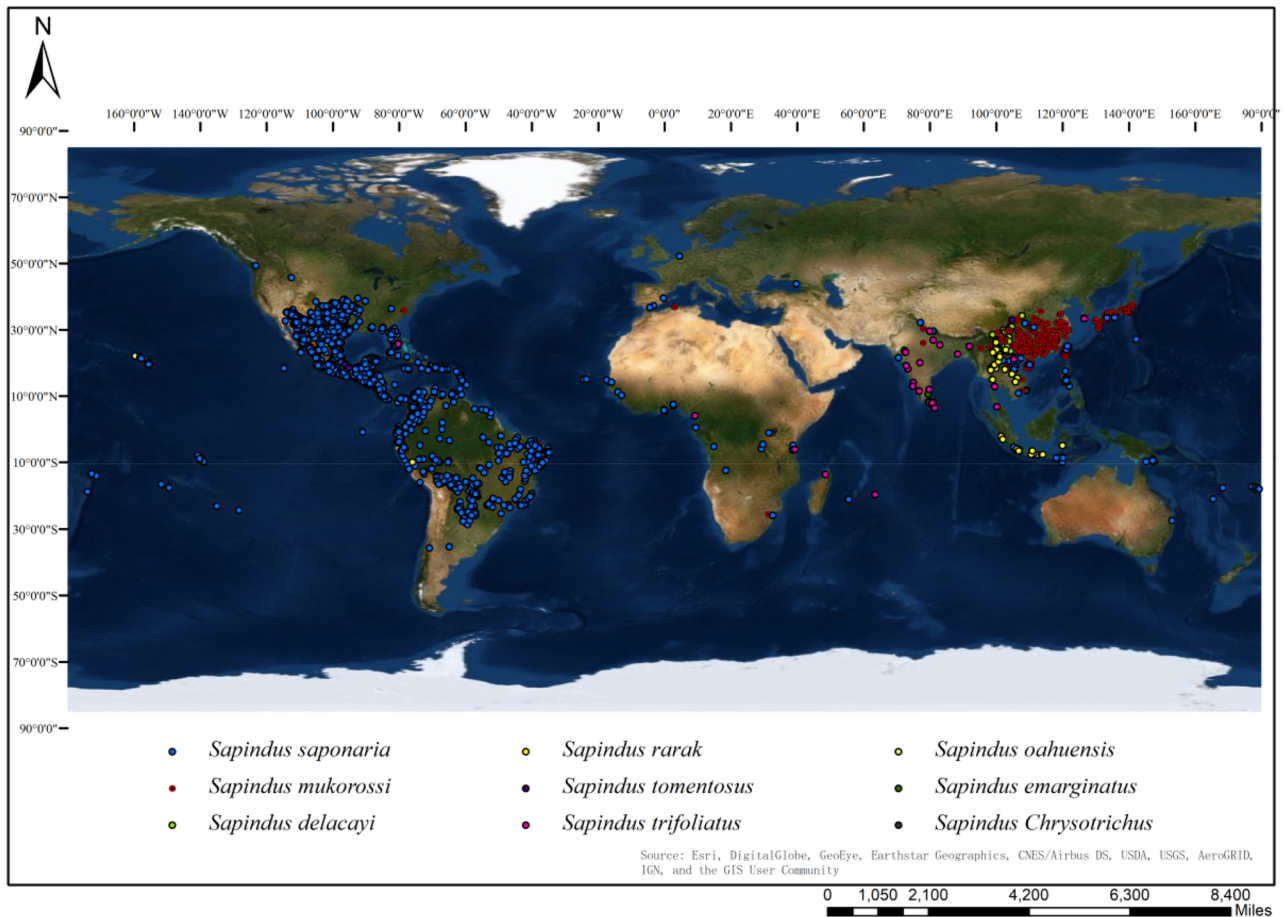
## Conclusions

Our modelling study showed that temperature, soil moisture, and soil pH may play an important role in determining the global distribution of suitable habitats for *Sapindus*. The ecological thresholds for critical environmental factors affecting *Sapindus* were the minimum temperature of the coldest month (0–20 °C), soil moisture (40–140 mm), mean temperature of the driest quarter (2–25 °C), mean temperature of the wettest quarter (19–28 °C), and soil pH (5.6–7.6). We found that the total suitable habitat area for *Sapindus* was  $6059.97 \times 10^4$  km<sup>2</sup>, which was unevenly distributed across six continents. High suitability areas were found to be concentrated in the Mississippi Plain, Florida Peninsula, and Mexico in North America; La Plata Plain in South America; and the North China Plain, Middle and Lower Yangtze Plain, Sichuan Basin, Yunnan–Guizhou Plateau, and Southern Himalayas in Asia. By integrating simulations from four future climate scenarios from 2020 to 2100, the size and distribution of suitable habitats for *Sapindus* were found to vary significantly across different continents depending on the climatic scenario. As greenhouse gas emissions increased (SSP126 to SSP585) and time passed (2021 to 2100), the overall area of suitable habitats contracted to an increasing degree, resulting from a significant contraction in lower latitudes and a slight expansion to higher latitudes. Suitable habitats for *Sapindus* in the northern hemisphere predominantly expanded, while in the southern hemisphere they contracted under the future climate change. Therefore, germplasm surveys and resource conservation should be prioritised in areas of the southern hemisphere that will likely become unsuitable for *Sapindus* in the future. In contrast, the Mississippi Plain, the Florida Peninsula, Mexico, Central America, the Middle and Lower Yangtze Plain, the Yungui Plateau, and central India can be used as a base for resource conservation, large-scale cultivation, and utilisation of *Sapindus* in the future.

## Materials and methods

**Design framework.** Based on global *Sapindus* occurrence data combined with historical environmental factors and future climate models, we explored the current and future potential habitats of *Sapindus* by utilising the MaxEnt (MaxEnt version 3.4.1) model and ArcGIS (ArcGIS 10.5). The specific process is shown in Fig. 4.





**Figure 5.** Global distribution of occurrence records for *Sapindus*. Created in ESRI ArcMap 10.5 (<https://support.esri.com/en/Products/Desktop/arcgis-desktop/arcmap>).

***Sapindus* occurrence records.** We collected 5674 worldwide *Sapindus* occurrence records from well-designed surveys, including: (1) 205 occurrence records from field germplasm surveys of *Sapindus* in China<sup>3</sup>, (2) 21 occurrence records from the Chinese National Plant Specimen Resource Center (CVH, <http://www.cvh.ac.cn/>), (3) 5425 occurrence records from the Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>)<sup>67</sup>, and (4) 23 occurrence records from the Chinese National Specimen Information Infrastructure (NSII, <http://www.nsii.org.cn/2017/>). There were 4505 *S. saponaria*, 893 *S. mukorossi*, 108 *S. delavayi* (Franch.) Radlk., 84 *S. rararak* DC., 50 *S. trifoliatus* L., 15 *S. oahuensis*, 12 *S. emarginatus* Vahl, 4 *S. tomentosus* Kurz, and 3 *S. chryso-trichus* germplasm. However, there were some spatial clusters of occurrence records, especially in southern North America and Southeast Asia. When these spatial clusters exist, models are often over-fitted in terms of environmental biases, and model performance values are inflated<sup>68,69</sup>. Therefore, we employed the Spatially Rarefy Occurrence Data tool in SDMtoolbox 2.0<sup>70</sup> and set the spatial interval at 10 km to eliminate spatial clusters of occurrences. After elimination, 2245 occurrence records were used in the model (Fig. 5).

**Environmental parameters.** There were 32 environmental factors used in this projection, including bioclimatic, topographical, UV-radiation, and soil parameters, to model the potential suitable habitat for *Sapindus* (Table S2). The bioclimatic factors for current climate conditions were extracted from the 2.5-min resolution historical climate (average for 1970–2000) database (<https://www.worldclim.org/data/worldclim21.html>). Topographical factors were obtained from the Harmonized World Soil Database v1.2 from the Food and Agriculture Organization of the United Nations (<http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/>). Soil factors were extracted from the Center for Sustainability and the Global Environment dataset (<https://nelson.wisc.edu/sage/>). UV-radiation variables were obtained from the glUV dataset (<https://www.ufz.de/gluv/>)<sup>71</sup>.

Future climate projections were extracted from the 2.5-min resolution Shared Socioeconomic Pathways (SSPs) scenarios available from the BCC-CSM2-MR global climate (2021–2040, 2041–2060, 2061–2080, and 2081–2100) database of the Coupled Model Intercomparison Projects 6 (CMIP6) (<https://www.worldclim.org/>), which is a project of the World Climate Research Programme (WCRP)'s Working Group of Coupled Modelling (WGCM). This included the SSP1-2.6 (ssp126), SSP2-4.5 (ssp245), SSP3-7.0 (ssp370), and SSP5-8.5 (ssp585) scenarios, all of which were considered possible future climates depending on the quantity of greenhouse gases emitted in the near future. All environmental factors were statistically resampled to a 2.5-min resolution using ArcGIS.

However, there were several environmental factors that may have exhibited collinearity. Strong collinearity between these factors may artificially inflate the accuracy of a model<sup>72</sup>. In this study, we applied the remove highly correlated variables tool in SDMtoolbox to avoid the potential problem of multicollinearity among environmental factors. Set 0.9 as the maximum correlation allowed value, we retained 21 environmental factors for subsequent modelling, respectively were Annual mean temperature (Bio1), Mean diurnal range (Bio2), Isothermality (Bio3), Max temperature of warmest month (Bio5), Min temperature of coldest month (Bio6), Temperature annual range (Bio7), Mean temperature of wettest quarter (Bio8), Mean temperature of driest quarter (Bio9), Mean temperature of warmest quarter (Bio10), Annual precipitation (Bio12), Precipitation seasonality (Bio15), Precipitation of wettest quarter (Bio16), Precipitation of driest quarter (Bio17), Precipitation of warmest quarter (Bio18), Precipitation of coldest quarter (Bio19), elevation (Elv), Soil Moisture (Sm), Soil Organic Carbon (Soc), Soil pH (Sph), Annual mean UVb (AmUV), Seasonality UVb (SeaUV).

**Model application.** The MaxEnt model utilises the maximum entropy principle, applying five different feature constraints (linear, product, hinge, quadratic, and threshold) to environmental variables to calculate the potential geographic distribution probability of a species<sup>53,73</sup>. We used the SDMtoolbox to run the MaxEnt model on the ArcGIS platform. We set 25% of the occurrence data as testing data and 75% of the occurrence data as training data, output format was cloglog, and applied percentile training presence as threshold rule. To determine the key environmental factors that drive *Sapindus* habitat distribution, we performed a jackknife permutation to rank the environmental factors and visualised the results in response curves. To calibrate and validate the robustness of the MaxEnt model, ture kil statistic (TSS)<sup>74</sup> and receiver operating characteristic curve (ROC curve) analysis were used, TSS can range from -1 to 1 and the area under the receiver operating curve (AUC) ranges from 0 to 1<sup>75,76</sup>. Model performance was classified as failing (0.5–0.6), poor (0.6–0.7), fair (0.7–0.8), good (0.8–0.9), and excellent (0.9–1) according to the AUC value<sup>77</sup>. TSS values closer to 1 indicated more successful models<sup>74</sup>.

We converted the continuous suitability score (0–1) of the MaxEnt model output into a habitat distribution visualisation using ArcGIS. We reclassified suitability into four classes: unsuitable habitat (<0.25), low suitability (0.25–0.50), suitable habitat (0.50–0.75), and high suitability (>0.75)<sup>78,79</sup>.

In the future distribution model, the future climate scenarios of four greenhouse gas emission models were used to simulate the global migration and change of suitable habitats for *Sapindus* from 2020 to 2100. Topographic and soil factors were set as stabilising variables in future models since topographic and soil factors are largely unaffected by climate change. In order to visualise the expansion and contraction of suitable habitats for *Sapindus* under different future climate scenarios, we created habitat suitability maps using ArcGIS and compared the current suitable habitats with those in each future climate scenario.

## Data availability

The datasets generated during and/or analysed during the current study are available in the Chinese National Plant Specimen Resource Center (CVH, <http://www.cvh.ac.cn/>), Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>), Chinese National Specimen Information Infrastructure (NSII, <http://www.nsii.org.cn/2017/>), WorldClim dataset (<https://www.worldclim.org/data/worldclim21.html>), Harmonized World Soil Database v1.2 from the Food and Agriculture Organization of the United Nations (<http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/>), Center for Sustainability and the Global Environment dataset (<https://nelson.wisc.edu/sage/>), and gLUV dataset (<https://www.ufz.de/gluv/>).

Received: 8 May 2021; Accepted: 6 September 2021

Published online: 05 October 2021

## References

- Liu, J. *et al.* Research progress in *Sapindus* L. germplasm resources. *World For. Res.* **30**, 15–21 (2017).
- Sun, C., Jia, L., Xi, B., Wang, L. & Weng, X. Natural variation in fatty acid composition of *Sapindus* spp. seed oils. *Ind. Crops Prod.* **102**, 97–104 (2017).
- Liu, J. *et al.* Variation in fruit and seed properties and comprehensive assessment of germplasm resources of the genus *Sapindus*. *Sci. Silva Sin.* **55**, 44–54 (2019).
- Xu, Y., Jia, L., Chen, Z. & Gao, Y. Advances on triterpenoid Saponin of *Sapindus mukorossi*. *Chem. Bull.* **081**, 1078–1088 (2018).
- Basu, A., Basu, S., Bandyopadhyay, S. & Chowdhury, R. Optimization of evaporative extraction of natural emulsifier cum surfactant from *Sapindus mukorossi*: Characterization and cost analysis. *Ind. Crops Prod.* **77**, 920–931 (2015).
- Mukhopadhyay, S., Hashim, M. A., Sahu, J. N., Yusoff, I. & Gupta, B. S. Comparison of a plant based natural surfactant with SDS for washing of As(V) from Fe rich soil. *Journal of Environmental Sciences* **25** (2013).
- Mukhopadhyay, S. *et al.* Ammonium-based deep eutectic solvents as novel soil washing agent for lead removal. *Chem. Eng. J.* **294**, 316–322 (2016).
- Mukherjee, S. *et al.* Optimization of pulp fibre removal by flotation using colloidal gas aphrons generated from a natural surfactant. *J. Taiwan Inst. Chem. Eng.* **53**, 15–21 (2015).
- Shinobu-Mesquita, C. *et al.* Cellular structural changes in candida albicans caused by the hydroalcoholic extract from *Sapindus saponaria* L. *Molecules* **20**, 9405–9418 (2015).
- Smulek, W. *et al.* Sapindus saponins' impact on hydrocarbon biodegradation by bacteria strains after short- and long-term contact with pollutant. *Colloids Surf. B* **142**, 207–213 (2016).
- Rodríguez-Hernández, D. *et al.* Highly potent anti-leishmanial derivatives of hederagenin, a triperpenoid from *Sapindus saponaria* L. *Eur. J. Med. Chem.* **124**, 153–159 (2016).
- Rodríguez-Hernández, D., Demuner, A. J., Barbosa, L. C. A., Csuk, R. & Heller, L. Hederagenin as a triterpene template for the development of new antitumor compounds. *Eur. J. Med. Chem.* **105**, 57–62 (2015).
- Singh, P. T. D. & Singh, M. M. Anti-Trichomonas activity of *Sapindus saponins*, a candidate for development as microbicidal contraceptive. *J. Antimicrob. Chemother.* **62**, 526–534 (2008).

14. Muntaha, S. T. & Khan, M. N. Natural surfactant extracted from *Sapindus mukurossi* as an eco-friendly alternate to synthetic surfactant: A dye surfactant interaction study. *J. Clean. Prod.* **93**, 145–150 (2015).
15. Sun, C., Jia, L., Ye, H. O., Gao, Y. & Weng, X. Geographic variation evaluating and correlation with fatty acid composition of economic characters of *Sapindus* spp. fruits. *J. Beijing For. Univ.* **12**, 73–83 (2016).
16. Barry, C. & Cox, P. D. *Biogeography: An Ecological and Evolutionary Approach* (Blackwell, 1980).
17. Stocker, T. F. et al. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of IPCC the Intergovernmental Panel on Climate Change*, 95–123. [http://www.ipcc.ch/publications\\_and\\_data/publications\\_ipcc\\_fourth\\_assessment\\_report\\_wg1\\_report\\_the\\_physical\\_science\\_basis.htm](http://www.ipcc.ch/publications_and_data/publications_ipcc_fourth_assessment_report_wg1_report_the_physical_science_basis.htm) (2014).
18. Hatfield, J. L. & Prueger, J. H. Temperature extremes: Effect on plant growth and development. *Weather Clim. Extremes* **10**, 4–10 (2015).
19. Ray, R., Gururaja, K. V. & Ramchandra, T. V. Predictive distribution modeling for rare Himalayan medicinal plant *Berberis aristata* DC. *J. Environ. Biol.* **32**, 725–730 (2011).
20. Lawler, J. J. Climate change adaptation strategies for resource management and conservation planning. *Ann. N. Y. Acad. Sci.* **1162**, 79–98 (2009).
21. Richard et al. Will plant movements keep up with climate change? *Trends Ecol. Evol.* (2013).
22. Araújo, M. B. & Peterson, A. T. Uses and misuses of bioclimatic envelope modeling. *Ecology* **93**, 1527–1539 (2012).
23. Elith, J. & Leathwick, J. R. Species distribution models: Ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Syst.* **40**, 677–697 (2009).
24. Zhu, G., Liu, G., Bu, W. & Gao, Y. Ecological niche modeling and its applications in biodiversity conservation. *Biodiv. Sci.* **1**, 94–102 (2013).
25. Stockwell, D. & Peters, D. P. The GARP modelling system: Problems and solutions to automated spatial prediction. *Int. J. Geogr. Inf. Sci.* **13**, 143–158 (1999).
26. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259 (2006).
27. Beaumont, L. J. et al. Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distribution. *Ecol. Model.* **186**, 251–270 (2005).
28. Liaw, A. & Wiener, M. Classification and regression by random forest. *R News* **23**, 1–10 (2002).
29. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–814 (2008).
30. Elith, J. et al. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **17**, 43–57 (2011).
31. Li, G., Du, S. & Wen, Z. Mapping the climatic suitable habitat of oriental arborvitae (*Platycladus orientalis*) for introduction and cultivation at a global scale. *Sci. Rep.* **6**, 30009 (2016).
32. Merow, C., Smith, M. J. & Silander, J. A. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography* **36**, 1058–1069 (2013).
33. Merow, C. et al. What do we gain from simplicity versus complexity in species distribution models?. *Ecography* **37**, 1267–1281 (2014).
34. Wang, W. et al. Assessment of potential habitat for firmiana danxiaensis, a plant species with extremely small populations in danxiahan national nature reserve based on maxent model. *Scientia Silvae Sinicae* (2019).
35. Guo, Y., Guo, J., Shen, X., Wang, G. & Wang, T. Predicting the bioclimatic habitat suitability of *Ginkgo biloba* L. in China with field-test validations. *Forests* <https://doi.org/10.3390/f10080705> (2019).
36. Huang, Z. et al. Geographic distribution and impacts of climate change on the suitable habitats of *Zingiber* species in China. *Ind. Crops Prod.* **138**, 111429 (2019).
37. Rong, Z. et al. Modeling the effect of climate change on the potential distribution of Qinghai Spruce (*Picea crassifolia* Kom.) in Qilian Mountains. *Forests* <https://doi.org/10.3390/f10010062> (2019).
38. Mohammadi, S., Ebrahimi, E., Moghadam, M. S. & Bosso, L. Modelling current and future potential distributions of two desert jerboas under climate change in Iran. *Ecol. Inform.* **52**, 7–13 (2019).
39. Ramos, R. S., Kumar, L., Shabani, F. & Picanco, M. C. Risk of spread of tomato yellow leaf curl virus (TYLCV) in tomato crops under various climate change scenarios. *Agric. Syst.* **173**, 524–535. <https://doi.org/10.1016/j.agsy.2019.03.020> (2019).
40. Sultana, S., Baumgartner, J. B., Dominiak, B. C., Royer, J. E. & Beaumont, L. J. Impacts of climate change on high priority fruit fly species in Australia. *PLoS ONE* **15**, e0213820 (2020).
41. Raffini, F. et al. From nucleotides to satellite imagery: Approaches to identify and manage the invasive pathogen *Xylella fastidiosa* and its insect vectors in Europe. *Sustainability* **12**, 4508 (2020).
42. Melo-Merino, S. M., Reyes-Bonilla, H. & Lira-Noriega, A. Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecol. Model.* <https://doi.org/10.1016/j.ecolmodel.2019.108837> (2020).
43. Sterne, T. K., Retchless, D., Allee, R. & Highfield, W. Predictive modelling of mesophotic habitats in the north-western Gulf of Mexico. *Aquat. Conserv. Mar. Freshw. Ecosyst.* <https://doi.org/10.1002/aqc.3281> (2020).
44. Convertino, M., Annis, A. & Nardi, F. Information-theoretic portfolio decision model for optimal flood management. *Environ. Model. Softw.* **119**, 258–274 (2019).
45. Ardestani, E. G. & Mokhtari, A. Modeling the lumpy skin disease risk probability in central Zagros Mountains of Iran. *Prev. Vet. Med.* **176**, 104887–104887. <https://doi.org/10.1016/j.prevetmed.2020.104887> (2020).
46. Hanafi-Bojd, A. A., Vatandoost, H. & Yaghoobi-Ershadi, M. R. Climate change and the risk of malaria transmission in Iran. *J. Med. Entomol.* **57**, 50–64. <https://doi.org/10.1093/jme/tjz131> (2020).
47. Zhang, L., Jing, Z., Li, Z., Liu, Y. & Fang, S. Predictive modeling of suitable habitats for *Cinnamomum Camphora* (L.) presl using maxent model under climate change in China. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph16173185> (2019).
48. Peng, L.-P. et al. Modelling environmentally suitable areas for the potential introduction and cultivation of the emerging oil crop *Paeonia ostii* in China. *Sci. Rep.* <https://doi.org/10.1038/s41598-019-39449-y> (2019).
49. Li, J., Fan, G. & He, Y. Predicting the current and future distribution of three Coptis herbs in China under climate change conditions, using the MaxEnt model and chemical analysis. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2019.134141> (2020).
50. Zhang, K., Zhang, Y. & Tao, J. Predicting the potential distribution of *Paeonia veitchii* (Paeoniaceae) in China by incorporating climate change into a maxent model. *Forests* <https://doi.org/10.3390/f10020190> (2019).
51. Sun, C. et al. Genetic structure and biogeographic divergence among *Sapindus* species: An inter-simple sequence repeat-based study of germplasm in China. *Ind. Crops Prod.* **118**, 1–10 (2018).
52. Mahar, K. S., Palni, L. M. S., Ranade, S. A., Pande, V. & Rana, T. S. Molecular analyses of genetic variation and phylogenetic relationship in Indian soap nut (*Sapindus* L.) and closely related taxa of the family Sapindaceae. *Meta Gene*, S2214540017300336 (2017).
53. Li, J., Chang, H., Liu, T. & Zhang, C. The potential geographical distribution of *Haloxylon* across Central Asia under climate change in the 21st century. *Agric. For. Meteorol.* **275**, 243–254. <https://doi.org/10.1016/j.agrformet.2019.05.027> (2019).
54. Prev y, J. S., Parker, L. E., Harrington, C. A., Lamb, C. T. & Proctor, M. F. Climate change shifts in habitat suitability and phenology of huckleberry (*Vaccinium membranaceum*). *Agric. For. Meteorol.* **280**, 107803 (2020).
55. Adeyemi, T. O., Ogunipe, O. T. & Olowokudejo, J. D. Species distribution modelling of family Sapindaceae in West Africa. *Int. J. Bot.* **8**, 45–49 (2012).

56. Sun, C. *et al.* Association of fruit and seed traits of sapindus mukorossi germplasm with environmental factors in Southern China. *Forests* **8**, 491 (2017).
57. Sun, C. *et al.* Genetic diversity and association analyses of fruit traits with microsatellite ISSRs in Sapindus. *J. For. Res.* **30**, 197–207 (2019).
58. Gao, Y. *et al.* Canopy characteristics and light distribution in *Sapindus mukorossi* Gaertn. are influenced by crown architecture manipulation in the hilly terrain of Southeast China. *Sci. Hortic.* **240**, 11–22 (2018).
59. Zhang, Y.-Q. *et al.* Spatio-temporal effects of canopy microclimate on fruit yield and quality of Sapindus mukorossi Gaertn. *Sci. Hortic.* **251**, 136–149 (2019).
60. Liu, J. *et al.* Potential distribution and ecological characteristics of genus *Sapindus* in China based on MaxEnt model. *Sci. Silva Sin.* **57**, 1–12. <https://doi.org/10.11707/j.1001-7488.20210501> (2021).
61. Cabral, A. L., Sales, J. D. F., Barbosa, K. F., Rodrigues, A. A. & Filho, S. C. V. Dormancy breakage and germination in *Sapindus saponaria* L. seeds as a function of temperature and germination substrate. *Semina* **40**, 3345–3358 (2019).
62. Jia, L. & Sun, C. Research progress of biodiesel tree *Sapindus mukorossi*. *J. China Agric. Univ.* **017**, 191–196 (2012).
63. Wei, X., Dai, T., Liu, S. & Jia, L. Effects of formula fertilization on leaf nutrient dynamics and yield of *Sapindus mukorossi* Gaertn. *J. Nanjing For. Univ.* **42**, 21–28 (2018).
64. Pal, A. K., Vaishnav, V., Meena, B., Pandey, N. & Rana, T. S. Adaptive fitness of *Sapindus emarginatus* Vahl populations towards future climatic regimes and the limiting factors of its distribution. *Sci. Rep.* **10**, 1–11 (2020).
65. Jayasinghe, S. L. & Kumar, L. Modeling the climate suitability of tea [*Camellia sinensis*(L.) O. Kuntze] in Sri Lanka in response to current and future climate change scenarios. *Agric. For. Meteorol.* **272–273**, 102–117 (2019).
66. He, X., Ning, X., Guo, Y. & Wei, H. Geographical distribution of *Xanthoceras sorbifolia* Bunge in China and predicting suitable area under the climate change scenario. *Res. Agric. Modern.* **40**, 138–146 (2019).
67. GBIF.org (13 October 2020) *GBIF Occurrence Download* <https://doi.org/10.15468/dl.4d9kye>.
68. Veloz, S. D. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *J. Biogeogr.* **36**, 2290–2299 (2010).
69. Boria, R. A., Olson, L. E., Goodman, S. M. & Anderson, R. P. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol. Model.* **275**, 73–77 (2014).
70. Brown, J. L. SDMtoolbox: A python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods Ecol. Evol.* **5**, 694–700. <https://doi.org/10.1111/2041-210x.12200> (2014).
71. Beckmann, M. *et al.* glUV: A global UV-B radiation data set for macroecological studies. *Methods Ecol. Evol.* **5**, 372 (2014).
72. Mao, J. F. & Wang, X. R. Distinct niche divergence characterizes the homoploid hybrid speciation of *Pinus densata* on the Tibetan Plateau. *Am. Nat.* **177**(4), 424–439 (2011).
73. Li, G., Xu, G., Guo, K. & Du, S. Mapping the global potential geographical distribution of black locust (*Robinia Pseudoacacia* L.) using herbarium data and a maximum entropy model. *Forests* **5**, 2773–2792 (2014).
74. Allouche, O., Tsoar, A. & Kadmon, R. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**, 1223–1232 (2006).
75. Elith, J. *et al.* Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151 (2006).
76. Phillips, S. J. Transferability, sample selection bias and background data in presence-only modelling: A response to Peterson *et al.* (2007). *Ecography* **31**, 272–278 (2008).
77. Araujo, M. B., Pearson, R. G., Thuiller, W. & Erhard, M. Validation of species–climate impact models under climate change. *Glob. Change Biol.* **11**, 1504–1513 (2005).
78. Ikhumhen, H. O., Li, T., Lu, S. & Matomela, N. Assessment of a novel data driven habitat suitability ranking approach for *Larus relictus* specie using remote sensing and GIS. *Ecol. Model.* **432**, 109221 (2020).
79. Bosso, L., De Conno, C. & Russo, D. Modelling the risk posed by the zebra mussel *Dreissena polymorpha*: Italy as a case study. *Environ. Manage.* **60**, 304–313 (2017).

## Acknowledgements

We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modelling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF.

## Author contributions

Conceptualization, J.L. and L.W.; methodology, J.L.; software, C.S.; validation, J.L., D.L., and Q.H.; formal analysis, J.L.; resources, X.W.; writing—original draft preparation, J.L.; writing—review and editing, B.X.; visualisation, Z.C.; supervision, L.J.; project administration, L.J.; funding acquisition, L.J. All authors have read and agreed to the published version of the manuscript.

## Funding

This research was funded by the Special Foundation for National Science and Technology Basic Research Program of China (Grant Number 2019FY100803) and the Fundamental Research Funds for the Central Universities (Grant Number 2019YC19).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-98389-8>.

**Correspondence** and requests for materials should be addressed to L.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021