Regional training workshop

Assessing the availability of tree seed sources for forest and landscape restoration

Coimbatore, India, 28 November to 2 December 2022

Introduction

The regional training workshop "Assessing the availability of tree seed sources for forest and landscape restoration" was organised in Coimbatore, India, from 28 November to 2 December 2022. The purpose of the workshop was to equip forestry experts in Asian countries to carry out spatial analyses on the availability of tree seed sources for native tree species. The specific learning objectives were to enable participants to:

- Model habitat suitability of tree species by tuning Maximum Entropy (MaxEnt) models
- Identify gaps in the availability of seed sources by overlaying species distributions, seed zone maps and locations of existing seed sources

The training was delivered using R Studio and QGIS which are both freely accessible software. The training was attended by 13 experts from five national research institutions in Bangladesh, India, Indonesia, the Philippines, and Laos. It was delivered by Dr Tobias Fremout of the Alliance of Bioversity International and CIAT, and hosted by the Institute of Forest Genetics and Tree Breeding (IFGTB) of India. The training was sponsored by the UK Darwin Initiative and the OneCGIAR Initiatives on Nature-positive Solutions and Sustainable Intensification in Mixed Farming Systems.

The training workshop was organised as part of the regional project "Strengthening collaborative tree seed supply systems for restoration in Asia" (2022-2024), led by the Alliance and funded by the UK Darwin Initiative. The project seeks to strengthen institutional and technical capacities to develop seed supply chains for native tree species in Bangladesh, India, Indonesia and the Philippines, so that forest and landscape restoration projects are linked with quality seed sources and local seed producers are linked with customers to support local livelihoods and sustainable forest management. The project will analyse gaps in the current availability of seed sources for pilot tree species in each country and support the identification of new seed sources and seed supply chains to fill the gaps.

The project is implemented as a collaboration of the following organisations:

- Alliance of Bioversity International and CIAT (lead)
- Bangladesh Forest Department
- Institute of Forest Genetics and Tree Breeding, India
- National Research and Innovation Agency (BRIN), Indonesia
- College of Forestry & Natural Resources, University of the Philippines Los Baños
- Royal Botanic Garden Edinburgh
- OECD Forest Seed and Plant Scheme

Workshop programme is given in Annex 1 and list of participants in Annex 2.
Workshop presentations are available from: https://bit.ly/3R97XCk
R-scripts are available upon request.

Day 1

Opening

A formal opening ceremony was organized on the morning of 28 November. The ceremony was graced by Director General of the Indian Council of Forestry Research and Education, **Shri Arun Singh Rawat**, IFS.

Figure 1. Director General of ICFRE lighting the lamp at the workshop opening ceremony

Theoretical background on species distribution modelling

Tobias Fremout provided theoretical background of species distribution modelling, referring to the use of mathematical models to predict the distribution of species across geographic space using environmental data (climate, soil, etc.). Distribution modeling is also known as environmental or ecological niche modelling (ENM) or habitat suitability modelling.

Distribution models are calibrated in a specific environmental space to obtain species' response functions to environmental variables. Calibrated models can then be applied to any set of climate data (other time periods or other parts of the world).

The extent to which modelled distributions coincide with real-world species distributions depends on the quality of environmental data, availability and reliability of species presence data, and complexity and accuracy of the models. Biotic interactions and dispersal limitations (e.g. between islands) increase the difference between the modeled niche and the realized niche in the field.

There are many modeling algorithms. The training course focused on Maximum Entropy (MaxEnt) models that will be tuned to optimize modeling results.

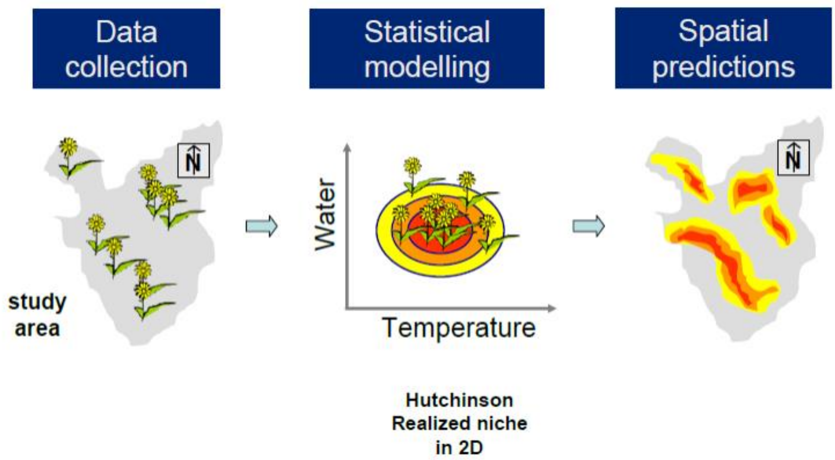


Figure 2. Overview of species distribution modeling. Source: Guisan & Zimmermann (2000)

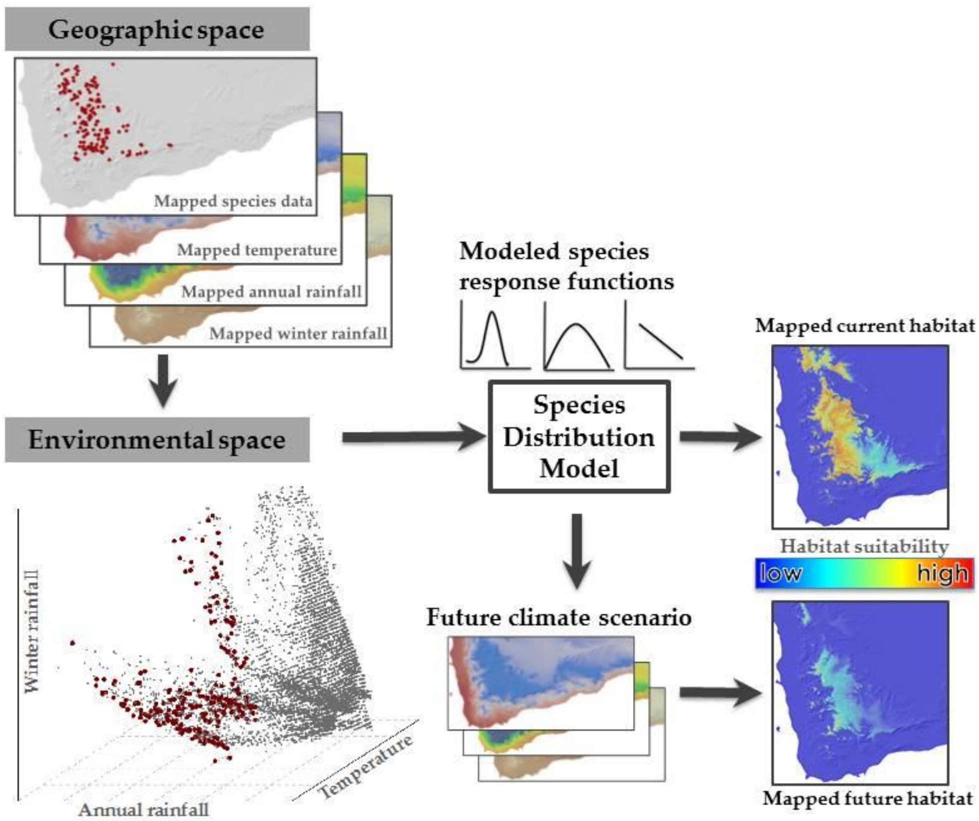


Figure 3. Model calibration and application to different climate conditions.

Discussion

Participants queried whether naturalized populations can be included in occurrence data. This depends on the objectives of modelling; if the objective is to find out where the species can be used in restoration, it likely makes sense to use records of naturalized populations. However, if the objective is to understand the species' natural distribution, one could either use all available occurrence points for model calibration and then mask out areas of non-native distribution on resulting maps; or use occurrence data of only natural populations from the outset. It may not always be possible to determine which populations are 'natural' and which naturalized, since humans have moved species around for millennia.

Participants also queried whether old herbaria records can be used as input data. The problem is that climate is changing and old herbaria records are not necessarily representative of species distributions under current climate. The base period used in climate models is usually 1961-2000, and it is recommended to exclude herbaria records older than this. At the same time, with the changing climate, the current base period is not representative of the current climate and we may need to soon transition to using a more recent base period in modeling.

Compilation and cleaning of species occurrence records

Participants retrieved occurrence data for their selected tree species from the Global Biodiversity Information Facility (GBIF) using R programming. They then cleaned the data to remove occurrences where the name of the province or state did not correspond to the coordinates. Many online sources have limited data available for Asian tree species, and so additional data sources such as herbaria, forest inventories and project data are important. Participants had compiled data from these and other similar data sources within their institutions and countries prior to the training.

Box 1. Requirements for occurrence data

Quantity: at least 30-50 points to obtain reliable models

Precision: should be at least as precise as the resolution of the environmental data used in the models (often 30 arcseconds, ca. 0.9 km at the equator)

Spread: should cover as much as possible of the environmental niche of a species

Spatial bias: less spatial bias is better, but some bias is unavoidable due to the practical limitations of data collection and sampling (e.g. it is common to have more occurrence data close to roads than further away). Effects of spatial bias can be reduced through spatial thinning, randomly removing points to achieve a more even distribution.

Preparing environmental data

The 19 bioclimatic variables from WorldClim were used as climate data. These are the most commonly used climatic variables in distribution modelling. Data from SoilGrids v1 was used for soil data. Participants experimented with visualising the retrieved environmental data using the QGIS software.

Box 2. Examples of sources of environmental data for species distribution modeling

Climate data:

- 19 bioclimatic variables from WorldClim, resolution of 30 arcsec – ca. 1 km (https://www.worldclim.org)
- CHELSA, resolution of 30 arcsec (https://chelsa-climate.org)

Soil data:

- SoilGrids – worldwide data at 250m resolution https://soilgrids.org

Topographic data:

- Many sources, for example: https://zenodo.org/record/1447210

Table 1. Environmental data used during the training workshop.

Climate: WorldClim	Soil: SoilGrids v1	Topographic variables
BIO1: Annual Mean Temperature	AWC: available water capacity	Negative openness
BIO2: Mean Diurnal Range	BLDFIE: bulk density	Positive openness
BIO3: Isothermality	CECSOL: cation exchange capacity	Slope
BIO4: Temperature Seasonality	CLYPPT: clay content	Topographic position index
BIO5: Max Temperature of Warmest Month	CRFVOL: coarse fragments volume fraction	Topographic wetness index
BIO6: Min Temperature of Coldest Month	ORCDRC: organic carbon content	
BIO7: Temperature Annual Range	PHIHOX: soil pH in H2O	
BIO8: Mean Temperature of Wettest Quarter	SLTPPT: silt content	
BIO9: Mean Temperature of Driest Quarter	SNDPPT: sand content	
BIO10: Mean Temperature of Warmest Quarter		
BIO11: Mean Temperature of Coldest Quarter		
BIO12: Annual Precipitation		
BIO13: Precipitation of Wettest Month		
BIO14: Precipitation of Driest Month		
BIO15: Precipitation Seasonality		
BIO16: Precipitation of Wettest Quarter		
BIO17: Precipitation of Driest Quarter		
BIO18: Precipitation of Warmest Quarter		
BIO19: Precipitation of Coldest Quarter		

Day 2

Distribution modelling

On the second day, participants proceeded to try out distribution modelling using MaxEnt algorithm[1]. MaxEnt compares the values of environmental data at presence (species occurrence)

[1] For information on the principles of MaxEnt modeling, see e.g.: Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. A statistical explanation of MaxEnt for ecologists. *Diversity and distributions*. 2011 Jan;17(1):43-57, and Merow C, Smith MJ, Silander Jr JA. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*. 2013 Oct;36(10):1058-69.

locations with the values at background locations (locations of a target group – group of species that are similar to the species being modelled). The method is designed to reduce spatial bias in presence data: presence locations of the target group are assumed to have a similar spatial bias than presence points of the modeled species, which compensates for the effects of spatial bias of presence points. For example, there is generally much more species occurrence data available in India's Western Ghats than in other parts of the country; however this reflects the number of studies conducted and not only suitability. MaxEnt is sometimes called 'presence-only' method. Most other modeling algorithms require absence records.

Resulting models were evaluated using the AUC statistic (area under the receiver operating characteristic curve, Figure 4) and the k-fold cross-validation method. AUC is the most common measure of predictive performance of species distribution models. The k-fold cross-validation method gives more accurate estimates of model performance than randomly partitioning data into training and testing data[2].

Participants then tested combinations of model parameters to arrive at the best-performing model for each study species. Model parameters that were varied were types of species response curves (feature classes) to environmental variables (Table 2), and *regularization multiplier* parameter which penalizes model complexity during model evaluation (less complex models receive better scores). Knowledge of species' biological characteristics can inform the selection of response curves.

Participants analysed the contribution of each environmental variable to the best-performing models, to assess which variables had the biggest impact on each species' modeled distribution. The contribution of each variable was determined by randomly permuting the values of that variable among the training points (both presence and background points), and measuring the resulting reduction in training AUC.

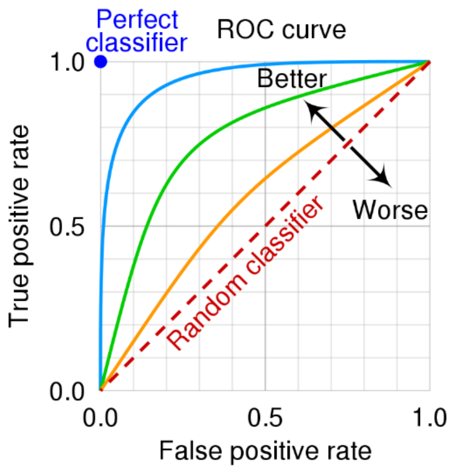


Figure 4. Using AUC (area under the receiver operating characteristic curve) to evaluate the performance of distribution models. AUC > 0.70 indicates a good model and AUC > 0.90 and excellent model.

[2] Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G. BlockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 2018, Jan 1:357798.

Table 2: Types of response curves for species response to the variation in environmental variables in MaxEnt models.

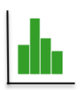
Feature type	Interpretation	Constraint	Shape
Linear	Continuous variable	The *mean* of each environmental variable at an unknown location should be close to the mean of that variable in known occurrence locations.	
Quadratic	Square of the variable	The *variance* of each environmental variable at an unknown location should be close to the variance of that variable in known occurrence locations.	
Product	Pairs of continuous variables – allows for interactions	The *co-variance* of two environmental variables at an unknown location should be close to the co-variance of those variables in known occurrence locations.	
Threshold	Conversion into binary response based on a threshold	The proportion of predicted occurrences with values above the threshold (binary response = 1) should be close to the proportion of known occurrences.	
Hinge	As threshold type, but response after the threshold (knot) is linear	The mean above the knot of each environmental variable at an unknown location should be close to the mean above the knot of that variable in known occurrence locations.	
Categorical	Categorical variable	The proportion of predicted occurrences in each category should be close to the proportion of observed occurrences in each category.	

Figure 5. Participants working in groups

Lastly, participants experimented with the threshold values to convert the suitability models into presence-absence models (Figure 5). Common methods for selecting the threshold values include:

- Maximizing specificity (true negative rate) + sensitivity (true positive rate)
- Setting a 10% omission threshold for presence locations

Almost any distribution model will result in some occurrence points falling outside of the modelled distribution area. The default 10% omission threshold can give models with too large distribution area in the presence-absence model. In this case, the threshold value can be increased, or other threshold methods can be used instead.

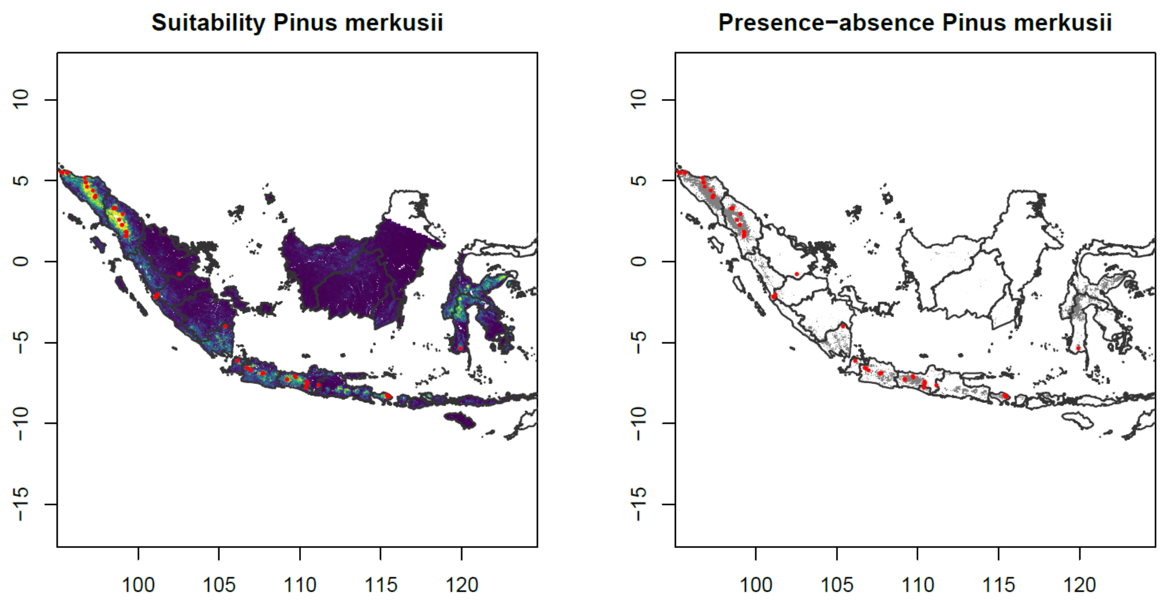


Figure 6: Suitability model (left) and presence-absence model (right) for *Pinus merkusii* in Indonesia. Predicted distribution area is shown in the presence-absence model in grey. Red dots indicate occurrence data.

Day 3

Gap analysis on seed source availability

Tobias Fremout provided brief background on seed zones. Seed zones are geographical areas within which environmental conditions are relatively similar, in which seeds (or other planting material) can be transferred with minimal risk of maladaptation or disrupting population genetic patterns. Seed zones should follow the spatial patterns of local adaptation within a species. Ideally, this information is obtained from common garden experiments (*provenance trials*). Alternative approaches include the use of genetic markers, or using only proxies for local adaptation, such as environmental data or ecoregions. When seed zones are based on the clustering of climate variables, they can be projected to future climatic conditions (*'dynamic' seed zones*)[3].

Seed sources should be available for each seed zone throughout the distribution of a species, so that suitably adapted seed and seedlings are available for forest and landscape restoration. Seed zones are a proxy for genetic variation, and so conserving seed sources in each zone also means conserving the genetic variation.

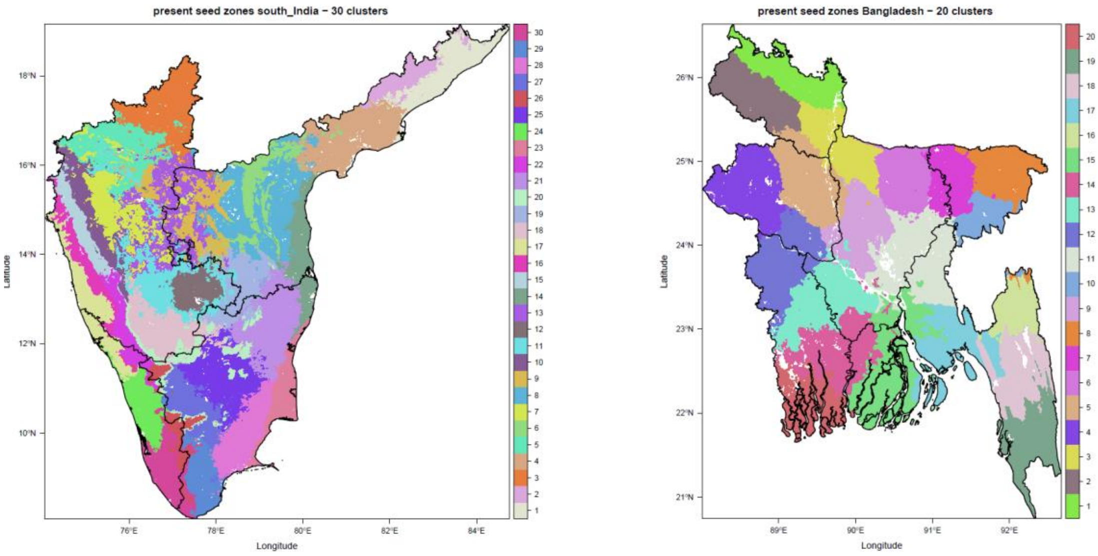


Figure 7: Seed zones for South India and Bangladesh, prepared using the methodology of Fremout et al. (2021).

Participants performed a gap analysis of seed source availability, by overlaying existing seed sources with the species' distribution maps and seed zones, and identifying seed zones that currently lack any known seed sources for the target species. These seed zones should be a priority for future surveys and activities to identify and establish seed sources.

Spatial analysis can help identify seed sources where they are lacking. Species modelled distribution can be overlaid, for example, with landcover maps to exclude non-forest areas, or with protected area maps or maps of community forests to identify potential seed sources. There are many free sources of land cover data, for example Copernicus (https://lcviewer.vito.be).

Discussion

Each country team gave a brief presentation of their results on distribution models and gaps in seed sources.

[3] Fremout T, Thomas E, Bocanegra-Gonzalez KT, Aguirre-Morales CA, Morillo-Paz AT, Atkinson R, Kettle C, González-M R, Alcazar-Caicedo C, Gonzalez MA, Gil-Tobon C. Dynamic seed zones to guide climate-smart seed sourcing for tropical dry forest restoration in Colombia. *Forest Ecology and Management.* 2021 Jun 15;490:119127.

Figure 8. Participants presenting their results. From top: Dr Rekha Warrier, Dr Vivi Yuskianti, Dr Md. Zahidur Rahman Miah and Dr Simone Vongkhamho

The day was concluded with a visit to the Institute of Forest Genetic Resources and Tree Breeding.

Figure 9. Participants visiting the nurseries and seed laboratory of the Institute of Forest Genetics and Tree Breeding. Bottom: Director of IFBTG C. Kunhikannan handing out certificates to participants, with trainer Tobias Fremout.

Day 4

A field trip was organised to the Silent Valley National Park in Kerala. The national park is one of the last undisturbed tracts of South Western Ghats mountain rain forests and tropical moist evergreen forest in India, and core area of the Nilgiri Biosphere Reserve recognised by UNESCO. The participants were guided by **Dr C. Kunhikannan**

Figure 9. Visiting the Silent Valley National Park

Annex 1: Programme

Day 1 – Monday 28 November

Time	Activity
8.30-9.30	Set up and testing Activity 1: Some theoretical background on species distribution models
9.30-10.15	Opening ceremony Group photo
10.15-10.45	Coffee break
10.45-12.30	Activity 2: Complementing the compiled data by downloading records from the Global Biodiversity Information Facility and other sources
12.30-14.00	Lunch
14.00-15.30	Activity 3: Cleaning of occurrence records
15.30-16.00	Coffee break
16.00-17.00	Activity 4: Preparing spatial environmental data for the region of interest
19.00	Workshop dinner

Day 2 – Tuesday 29 November

Time	Activity
8.30-10.00	Activity 1: Modelling the potential distribution (habitat suitability) of the selected tree species by tuning MaxEnt models
10.00-11.00	Coffee break
11.00-12.30	Activity 1: Modelling the potential distribution (cont'd)
12.30-14.00	Lunch
14.00-15.30	Activity 1: Modelling the potential distribution (cont'd)
15.30-16.00	Coffee break
16.00-17.00	Activity 2: Some theoretical background on seed zones and seed sources

Day 3 – Wednesday 30 November

Time	Activity
8.30-10.00	Activity 1: Identify gaps in seed sources and potential areas to identify new seed sources by overlaying species distributions, seed zone maps, land cover maps and locations of existing seed sources.
10.00-11.00	Coffee break
11.00-12.30	Activity 1: Identify gaps in seed sources (cont'd) and potential areas to identify new seed sources (cont'd)
12.30-14.00	Lunch
14.00-15.30	Activity 1: Identify gaps in seed sources (cont'd) and potential areas to identify new seed sources (cont'd)
15.30-16.00	Coffee break
16.00-16.45	Presentation of results by country teams
16.45-17.00	Reflections and closing

Day 4: Field trip

Time	Activity
8.00-10.30	Travel to the Silent Valley National Park
10.30-11.00	Registration and comfort break
11.00-13.00	Travel to the core area of the park and trek to a river side

Time	Activity
13.00-14.00	Lunch
14.00-16.00	Visiting the park
16.00-20.00	Return to Coimbatore

Day 5: Project planning meeting (national project coordinators only; optional for others)

Objectives

- Review progress with Year 1 activities
- Share good practices, experiences and ideas on methods and approaches for strengthening capacities in species data management and seed sourcing
- Prepare detailed country- and project-level workplans for 2023

Time	Activity
8.30-8.50	Good morning & Recap of the workshop
8.50-10.00	**Gap analysis: next steps to finalisation (Output 1)** Group discussion ('50) - Action needs to finalise the gap analysis for ≥5 species per country - Validation of results with key stakeholders: who and how? - How to publish and publicize the results to support uptake? - Optional: how to expand the analysis to new species Plenary ('20)
10.00-10.30	Coffee break
10.30-12.15	**Assessing current capacities and constraints of FLR implementers in seed sourcing for FLR (Output 3)** Presentation: Examples of capacity assessments on seed sourcing – *Riina* ('15) Group discussion ('45) - Identify stakeholders in project landscapes for capacity needs assessment - Plan process and timeline for surveys and interviews (Jan-March 2023) - Plan format and timeline for organising a training (Apr-June 2023) Plenary ('30)
12.15-13.30	Lunch
13.30-15.00	**Evaluating and improving existing databases on seed sources (Output 2)** Good practices in managing data on seed sources ('25) - *tbc* Group discussion ('45) - Revisit known knowledge gaps and challenges with data management, and knowledge needs to identify improvements - Identify stakeholders who need to be involved and how to engage them - Plan process for evaluating and improving databases Plenary ('20)
15.00-15.30	Coffee break
15.30-16.30	**Workplans and budgets** Group discussion ('20) - Compile actions from previous sessions as a workplan and budget Plenary ('40) - Brief presentations of workplans - Project-level activities
16.30-16.45	Closing and farewell

Annex 2: Participants

Training participants
Ambar Dwi Suseno, Directorate of Forest Tree Seed, Ministry of Environment and Forestry, Indonesia
R. Anandalakshmi, Institute of Forest Genetics and Tree Breeding, India
R. Archana, Institute of Forest Genetics and Tree Breeding, India
Bayu Arief Pratama, National Research and Innovation Agency (BRIN), Indonesia
Cristino S. Tiburan, College of Forestry & Natural Resources, University of the Philippines Los Baños
Dennis M Gilbero, Mindanao Tree Seed Center, Ecosystems Research and Development Bureau, Philippines
Enrique Tolentino jr., College of Forestry & Natural Resources, University of the Philippines Los Baños
Vivi Yuskianti, National Research and Innovation Agency (BRIN), Indonesia
Mahmudah Roksena Sultana, Bangladesh Forest Department
Md. Monsur Alam, Ministry of Environment, Forest and Climate Change
Md. Zahidur Rahman Miah, National Botanical Garden, Bangladesh
Rekha Warrier, Institute of Forest Genetics and Tree Breeding, India
Simone Vongkhamho, Forest Research Center, Lao PDR
Smitha Krishnan, Alliance of Bioversity International and CIAT

Trainer
Tobias Fremout, Alliance of Bioversity International and CIAT

Resource persons
Peter Wilkie, Royal Botanic Garden Edinburgh, UK (online for the project planning day)
Mirko Liesebach, Thünen Institute of Forest Genetics, Germany

Organisers
Riina Jalonen, Alliance of Bioversity International and CIAT, Malaysia
Rekha Warrier, Institute of Forest Genetics and Tree Breeding, India