Title: Harmonising distributed tree inventory datasets across India can fill critical gaps in tropical ecology

Authors:

Krishna Anujan [1,2,3]* (ORCiD: 0000-0003-3604-5895)
Neha Mohanbabu [1,4] (ORCiD: 0000-0002-6557-131X)
Abhishek Gopal [1,5,6] (ORCiD: 0000-0002-3020-4122)
Akshay Surendra [7,8] (ORCiD: 0000-0003-2719-7432)
Aparna Krishnan [1,9] (ORCiD: 0000-0001-9781-1157)
Ankitha Jayanth [1,10] (ORCiD: 0009-0003-3377-2605)
Tanaya Nair [1,11] (ORCiD: 0000-0003-2622-8612)
Shasank Ongole [1,12] (ORCiD: 0000-0001-6635-7027)
Mahesh Sankaran [9] (ORCiD: 0000-0002-1661-6542)

Affiliations:

[1] India Tree Inventory Network, National Centre for Biological Sciences, TIFR, Bellary Road, Bengaluru 560065, India
[2] Smithsonian National Zoo and Conservation Biology Institute, 1500 Remount Road, Front Royal, Virginia 22303, USA
[3] Smithsonian Environmental Research Centre, Edgewater, Maryland, USA
[4] University of Minnesota, St Paul, Minnesota, USA
[5] CSIR-Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad, India
[6] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India
[7] Lewis B. Cullman Fellow, New York Botanical Garden, Bronx, New York, USA
[8] School of the Environment, Yale University, New Haven, Connecticut, USA
[9] National Centre for Biological Sciences, TIFR, Bellary Road, Bengaluru 560065, India
[10] University of South Bohemia, Ceske Budejovice 37005 , Czech Republic
[11] Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, UK [12] Plant Ecology, University of Bayreuth, Universitaetsstrasse 30, 95447, Bayreuth, Germany

*corresponding author: krishna.anujan@gmail.com

Abstract

1. Global analyses of tree diversity and function are strongly biased geographically, with poor representation from forests of the Indian subcontinent. Even though data from India - representing two-thirds of the subcontinent and spanning a wide range of tree-based biomes - exists, a barrier to syntheses is the absence of accessible and standardised data. Further, with increasing human footprint across ecosystems, data from Indian landscapes, with their long history of human-nature interactions is a key link to understand the future of tropical forested landscapes. Given the long history of human-nature interactions and high human footprint in the region, accessible and standardized data from the subcontinent can enable understanding the future of tropical landscapes under increasing human footprints globally.

2. Combining literature searches with manual data retrieval, we assembled the INdia Tree Inventory dataset, INvenTree. INvenTree is the largest meta-dataset of peer-reviewed publications (n = 465) from 1991-2023 on geolocated plot-based tree inventories of multispecies communities from Indian ecosystems, in aggregate covering 4653.64 ha and all of its woody biomes.

3. Using INvenTree, we show extensive sampling across tropical moist and dry forests, the dominant ecosystem types in the country. However, most studies have low sampling effort (median sampled area = 2 ha) and data across studies is not openly accessible (73.33 % of studies representing 83.43% of the sampled area), potentially hindering inclusion into regional or global syntheses.

4. Significantly, we show majority authorship from within the country; 82.8% of corresponding authors were from India and 73.33% of the studies had all authors affiliated with Indian institutions. We also identify ecological and conservation sampling priority regions based on forest cover and forest loss and set a blueprint for future sampling efforts in the country.

5. Based on extensive Indian scholarship in forest ecology showcased through the INvenTree dataset, we see opportunity for regional collaboration to create scientific inferences that are larger and scalable, while prioritising data and knowledge equity. Harmonising these existing datasets, and synthesising historic and grey literature, will contribute enormously to understanding the human dimension of tropical ecology as well as informing regional management and conservation.

KEYWORDS

tree inventory; India; macroecology; bibliometrics

Introduction

Distributed tree inventories, typically collected using standardised methodologies, are powerful tools for addressing fundamental macro- and micro- ecological questions. Tree inventory information is a crucial baseline to understand ecological communities and their susceptibility and resilience in the face of global change including advancing our understanding of forest structure and dynamics, carbon stocks and fluxes, and improving predictions under climate change scenarios (Anderson-Teixeira et al., 2015, 2015, 2018; Cazzolla Gatti et al., 2022; ForestPlots.net et al., 2021, 2021; Jung et al., 2021; Malhi et al., 2002; Poulter et al., 2019). However, the global distribution of these tree inventory datasets is geographically biased, leading to highly variable and weakly constrained predictions from data-poor regions, many of which overlap deforestation and climate-change hotspots (Carvalho et al., 2023).

Continental biases in data pose a barrier to truly global ecology, both because of ecological and anthropic differences among the global tropics. Firstly, tropical biomes show continental scale differences in species richness, stem density and carbon stocks, because of differences in earth history and biogeography (ForestPlots.net et al., 2021; Hagen et al., 2021; Sullivan et al., 2017). Secondly, there is growing recognition of historical and ongoing human influence in shaping natural landscapes, influencing patterns, processes and functioning (Barlow et al., 2012; Fraser et al., 2024). This human influence on ecosystems has differences across continents and biomes, with some tropical biomes having only <1% under low human influence (Riggio et al., 2020; Venter et al., 2016). Yet, global analyses using tree inventories draw on sparse and biased data from the tropics, hindering truly "global" ecological knowledge (Martin et al., 2012; Stocks et al., 2008). Global syntheses use publicly available data (e.g. hosted at GBIF or Dryad) and aggregate national scale data (e.g. US Forest Inventory and Analysis), inadvertently reflecting and propagating the biases within these datasets. At the global scale, collection and compilation of biodiversity data is biased towards countries with high domestic product, high proportions of English speakers and high levels of security (Amano & Sutherland, 2013; Hughes et al., 2021). Temperate biomes are oversampled, while tropical biomes are understudied, and the Neotropics are studied more than the Paleotropics (Martin et al., 2012; Stocks et al., 2008). For a robust understanding of tropical ecology in the Anthropocene, ecological understanding must span variation on both ecological and human dimensions.

The Indian subcontinent is a pronounced gap in most global datasets and analyses, occluding a robust understanding of both ecological and human dimension of tropical ecology. The Indian subcontinent is one of the oldest and most biodiverse regions of tropical Asia, with a long human history and present-day footprint. Human settlement here has been recorded as early as 700,000 to 400,000 years before present (Briggs, 2003; Gadgil & Thapar, 1990; Mani, 1974). Among the global tropics, this region ranks highest in present-day human footprint on ecosystems, on par with highly built areas of the industrialised temperate regions (Venter et al., 2016). To this day, these ecosystems continue to support high human densities and human use,

for timber, fuelwood, fodder and non-timber forest products (Karanth & DeFries, 2010; Shahabuddin & Prasad, 2004-04/2004-06). Despite this unique combination of bioclimatic, biogeographic and human forces in shaping present-day natural landscapes, the subcontinent continues to be excluded from macroecological syntheses, including estimates of tree density, forest age and species richness (e.g. Cazzolla Gatti et al., 2022; Crowther et al., 2015; Poulter et al., 2019). These geographic gaps could be due to a combination of factors including gaps in data collection within the subcontinent (e.g. across biomes or regions of interest) or gaps in data reusability (e.g. data stored in repositories are missing data or metadata components).

Within the Indian subcontinent, data availability and accessibility could be affected not only by sampling constraints, but by legacies of scientific inequalities, further hindering collaboration and synthesis. Stemming from colonial legacies, tropical ecology research and publishing has been has been dominated by researchers from high-income countries, leading to "parachute science" (Maas et al., 2021; Miller et al., 2023; Nuñez et al., 2021; Odeny & Bosurgi, 2022; Raja et al., 2021; Stocks et al., 2008). Moreover, despite ecological research increasingly moving to a big-data and collaborative discipline, its costs and benefits continue to be unevenly distributed (Farley et al., 2018; Maas et al., 2021). This is especially stark for synthesis and meta-analyses requiring quantitative skills and no new data collections (Mori et al., 2015; Stocks et al., 2008). As voiced by researchers in the Neotropics and in Africa, researchers based in the Global South experience multistep challenges in the scientific publishing pipeline - lack of funding, language barriers, inequalities at the peer-review stage and inability to pay high open access publishing costs - as well as transient and transactional partnerships with Global North researchers (Asase et al., 2022; Kwon, 2022; Mekonnen et al., 2021; Ramírez-Castañeda, 2020; Smith et al., 2023). The Indian subcontinent, with legacies of its recent colonial past reflected in society, including in ecology and forestry practices, is likely affected by these same issues. A large and diverse region such as the Indian subcontinent is prone to parachute science by Global North researchers into the region, as well as parachute science within the region: researchers with more resources may extract data and knowledge from distant, often resource-poor regions within the subcontinent. Such academic inequalities could disincentivise data sharing, especially in the absence of clear incentives (De Lima et al., 2022). Since barriers to data access have strong social dimensions, it is important to carefully examine the geographic distribution of the researchers, biomes and publication patterns, to understand the community of data owners and assess collaboration needs.

As the largest country in the Indian subcontinent, and representative of its diversity and human footprint, data collation from India can further ecological understanding for the region, beyond biodiversity assessment. India, currently the most populous country in the world, continues to have large forest cover and high species diversity under four biodiversity hotspots (Mittermeier et al., 2004; Mugal et al., 2023; Myers et al., 2000). Ecosystems in India cover 10 biomes as proposed by Dinerstein et al. (2017), covering all biomes of South Asia. Moreover, ecosystems in India are human-dominated; they continue to support high human densities along with high

biodiversity to the extent that 85% of high priority conservation areas in India are outside of protected areas and embedded in wider human-nature landscapes (Karanth & DeFries, 2010; Srivathsa et al., 2023). India, therefore, provides a unique "model" to understand human-dominated tropical landscapes as the human footprint on natural landscapes across the world continues to grow (Karanth & DeFries, 2010; Roberts et al., 2021). Further, Indian scholarship in forest ecology is extensive, with a growing number of research programmes and investment, including over 250 long-term monitoring efforts (Tiruvaimozhi et al., 2024). Given its large human capital, collaborative frameworks from India can also set benchmarks for inclusive tropical ecology. Here, we assembled INvenTree, the largest meta-dataset of geolocated plot-based tree inventories from India and assessed (i) sampling patterns, (ii) geographic gaps in sampling and (ii) authorship patterns in tree community research in India to aid future syntheses.

Materials and Methods

Study area

The Indian subcontinent is one of the oldest regions in tropical Asia (Briggs, 2003; Mani, 1974). Its complex geo-climatic and biogeographic history has resulted in a rich variety of biomes with high diversity and endemism of flora and fauna (Fig 1). Within the subcontinent, India, 6˚ 440 and 35˚ 300 N latitude and 68˚ 70 and 97˚ 250 E longitude, comprises the largest landmass, with four global biodiversity hotspots occurring within mainland India (Mittermeier et al., 2004; Myers et al., 2000). It comprises ten biogeographic subdivisions based on phytogeography (Rodgers & Panwar, 1988), and five major broad forest types are commonly recognised: tropical moist forest, tropical dry forest, subtropical forest, temperate and sub-alpine forest (Champion & Seth, 1968; Reddy et al., 2015) (Fig S1). Among these, the tropical dry forest is the most dominant vegetation type in terms of land cover (Reddy et al., 2015). The current distribution of flora within the mainland is predominantly shaped by the interplay of past and current climatic events and biogeography (Mani, 1974). Seasonal rainfall driven by the Himalayan orogeny is one of the major drivers of current vegetation patterns with most of the subcontinent strongly governed by monsoonal rainfall. For ease of comparison with global studies we used the classification proposed by Dinerstein et al. (2017), which resulted in classification of the Indian mainland and the Andaman and Nicobar Islands into ten biomes with several ecoregions nested under these (Fig 1, Fig S2).

The Indian subcontinent has a long history of human populations with the earliest human settlement recorded between 700,000 to 400,000 years before present (Gadgil & Thapar, 1990). Human population spread throughout the subcontinent at low densities in the terminal pleistocene (20000 - 10000 years BP) and by 4000 years BP, agricultural settlements had covered the entire subcontinent. India is currently the most populous country in the world, with a population of 1.4 billion which is projected to continue rising in the coming decades [World

Population Prospects]. To this day, forested landscapes in India also house large densities of human populations with forest- or other natural-resource-dependent livelihoods.

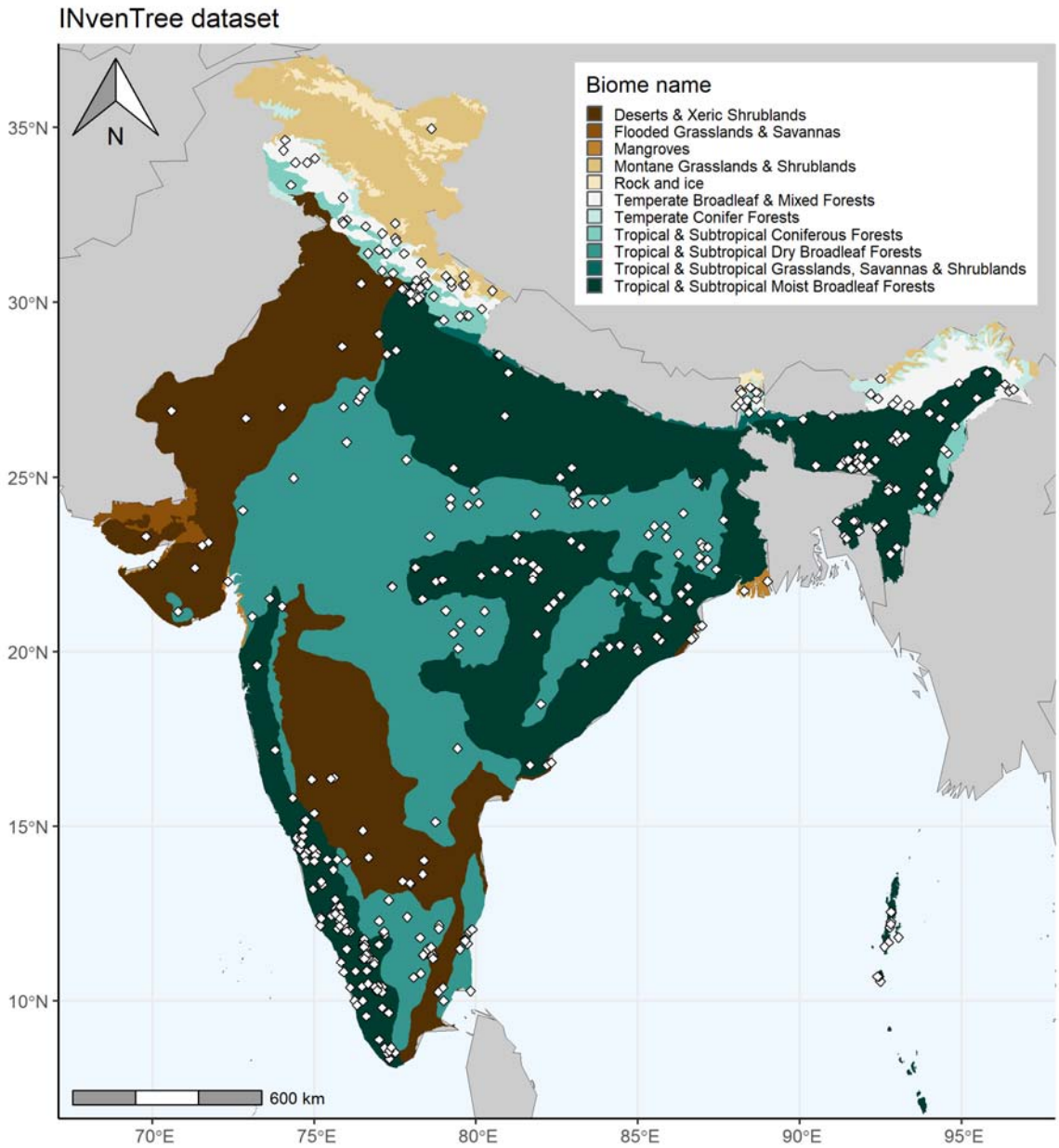

Figure 1: INvenTree dataset. *The location of study sites in papers that form part of the INvenTree dataset in relation to the 10 vegetated biomes (and Rock and Ice) in India as per the Dinerstein et al. (2017) dataset. Each point corresponds to the centroid of the sampled location reported in each study.*

ASSEMBLING THE INVENTREE DATASET

We created a meta-dataset of plot-based tree inventory studies on multispecies communities from Indian forested ecosystems from 1991 to 2023. We used both Google Scholar (last search in November 2021) and Web of Science (last search in February 2023) to include a wide range of journals as well as book chapters that publish forestry research from India. The Google Scholar search terms included (Tree diversity OR plot OR dbh) AND (India) and was later updated to exclude other countries where "India" or "Indians" may have been used in other contexts (e.g.,other countries in the Indian sub-continent such as Pakistan, Nepal, Myanmar etc., American Indian reservations). We used a broader list of search terms on Web of Science with the keywords AB=((tree diversity OR forest structure OR (tree* AND biomass) OR (forest* AND biomass) OR carbon stock OR vegetation survey OR vegetation sampling OR (tree* AND plot) OR (forest AND plot) OR (measur* AND tree) OR (checklist AND flora) OR (checklist AND tree) OR (checklist AND plant*)) AND India). We supplemented this list with an additional set of papers we were aware of but weren't present in the Google Scholar or Web of Science results (e.g. global studies that report on datasets from India but are not captured by the keywords), and with papers that we found through snowball sampling the known papers' references. We merged the lists from all three sources and removed duplicates, resulting in 3758 entries. We then manually sorted the list based on the title and the abstract to remove papers that were not related to forests in India, resulting in 657 potentially relevant papers. We accessed shortlisted papers and screened them based on the following criteria– (1) the study involved collection of plot level or checklist data of trees, (2) the study sites were natural multi-species stands of trees from inland woody ecosystems or coastal forests and mangroves. We excluded papers focussing on seedlings, single species studies and those based in managed or urban environments. Based on these criteria we identified **465 papers that formed the INvenTree dataset**.

From each paper in the INvenTree dataset, we extracted data on i) geographic location of study ii) area sampled iii) author affiliations and iv) data accessibility. i) We recorded the latitude and longitude of the study area reported either from the text or from a map provided. For studies with more than one plot, where separate locations were reported for each, we estimated the centroid of the study area. ii) To calculate the total area sampled in each study, we recorded plot shape, plot dimensions and number of plots, for all plot-based studies. Since we were focussed on exhaustive inventories, for studies that sampled subplots within a larger plot (e.g. five 20 m x 20 m plots sampled within a 1 ha area), we recorded the total area of the subplots which were exhaustively sampled (5x 400 m^2) instead of the reported study area (1 ha). Further, for studies that inventoried plots across natural and managed communities or plantations, we only counted plots that were in natural multi-species communities. (iii) We verified (for Web of Science search) or manually added information on author affiliations and addresses. (iv) To assess the accessibility of the full paper and data, we recorded whether the papers were freely available (open access) on the publisher's webpage and whether individual level or species level summaries of the tree data were freely accessible in the paper, in appendices, or in open access data sharing websites like FigShare, Dryad or Zenodo. Finally, we identified papers that used the same datasets - either large scale datasets or long-term datasets. We excluded large-scale

datasets from analysis of biome-distributions as they covered several biomes; for long-term datasets we included only the first published paper. For analyses of authorship, we used all published papers including large-scale and repeat datasets.

Analysing data accessibility and reusability

We first examined the number of studies, total area and percent area sampled within each biome. We also examined the open access status and the data access status of studies across biomes. Although some of our sites fell outside of biomes expected to have tree cover (e.g. in Montane grasslands and shrublands), we included them if they included tree inventories, as this information is likely to be more accurate than biome classifications. To further understand the practical constraints on the reuse of these data, we assessed each dataset on two axes: the data resolution and reusability. We considered these axes independently because for full reusability for synthesis, data should be available at the finest resolution, stem level data, but some synthesis might be possible with aggregate data as well. For each study that had reported data beyond broad summaries, we accessed the data from external repositories or from the paper. We used the FAIR framework for scientific data management (Wilkinson et al., 2016) to assess the reusability status of dataset in INvenTree, by categorising datasets in each study based on whether they are Findable, Accessible, Interoperable and Reusable. We categorised these datasets on the FAIR-ness axis, for three different data resolutions - plot-level, species-level and stem-level data. We defined plot-level data as data aggregated at the plot level, where plot is the smallest spatial unit of sampling, species-level data as data aggregated at the species level, whether within or across plots, and stem-level data as data reported at the level of data collection - individual stems. For detailed methods, refer to the supplementary material.

Analysing geographic patterns in sampling

We assessed sampling gaps in India in two different ways; we compared the spread of our data in relation to a) tree cover (as a proxy of ecological sampling priority) and b) its loss in India (conservation sampling priority). Indian forested ecosystems have a large range of natural tree cover, spanning closed canopy forests to open natural ecosystems. For robust ecological understanding, sampling should span this range of variation in tree cover. We defined ecological sampling priority as undersampled areas across the tree cover gradient. We also defined conservation priority sampling areas as areas of high tree cover loss but low sampling effort. We used the Hansen et al. (2013) dataset of tree cover in the year 2000 (within the period of interest) and tree cover loss from 2000-2021 cropped to the limits of India using Google Earth Engine (code here - https://code.earthengine.google.com/1f887c3af348300241dfc60e3da21cff). For these 30 m x 30 m pixels, we followed Friedl & Sulla-Menashe (2022)'s definition of forested pixels as >30% forest cover. This would include woody savannas (30-60% cover) but exclude more open and grassy savannas (10-30%).

India has 594 districts as of the 2011 census with a median area of 4077.48 km^2 ranging from 15.7 km^2 to 47858 km^2 area. We chose district as our sampling unit as the sampling scale of most of the studies in the INvenTree dataset was contained within a district. Moreover, ecological heterogeneity is often small enough within a district and summaries at this scale are meaningful for further use by multiple stakeholders (Srivathsa et al., 2023). We used an administrative map of districts from [GADM] and extracted the tree cover (number of forested pixels) and cover loss (number of pixels experiencing loss) for each district using the *extract* function from the *terra* package (Hijmans, 2022b). For each district, we then calculated our variables of interest - % tree cover (percentage of district area with tree cover) and % forest loss (number of pixels that experienced forest loss divided by number of forested pixels). We chose to analyse % forest loss instead of absolute forest loss since there is a large distribution of forest cover across districts and absolute loss would be somewhat correlated to forest cover (for analysis of absolute forest loss, refer to the supplementary appendix). We then summarised the proportion of area sampled in each district as the sum of the total area sampled in each study that fell within its boundaries divided by the district area. We limited the analysis to districts that have non-zero forest cover (n=341) and excluded all points that fell outside these even though they may have reported on tree-based biomes, because of high uncertainties.

To analyse the spatial distribution of data, we assigned each district into one of three categories (low, medium, high) along each variable assessed - % tree cover, % forest loss and sampling effort. For %tree cover, we split districts as three quantiles and scored them as 1-3, i.e. districts that fell in the lowest 1/3rd of % tree cover values was scored as 1. Many districts had zero tree cover loss (n=115 out of 341) or zero sampling effort (n=205) and so, it was not possible to split the districts into three evenly sized groups based on the values of these variables. Moreover, zero sampling and zero loss are qualitatively different from low sampling and need to be highlighted. We, therefore, assigned all districts with zero tree cover loss as category 1, and split the remaining districts evenly between categories 2 and 3. We used a similar logic for sampling effort, but since our aim was to highlight research gaps, for sampled effort, we assigned a reverse scale. All districts with no sampling were categorised as the highest "sampling gap" and assigned category 3. The remaining districts were split evenly between categories 1 and 2. In our scale, a district falling in the highest quantile of tree cover and no sampling effort would be scored (3,3) to highlight high priority ecological sampling. Similarly, a district with high loss and no sampling effort would be scored (3,3) to highlight high priority conservation sampling. We visualised spatial overlaps in tree cover or tree cover loss and sampling gaps using bivariate chloropleth maps. We repeated these steps for sampling gaps across ecoregions. For detailed methods and data distributions, refer to the supplementary material.

Analysing author affiliations and parachute science

We considered the full INvenTree dataset (n = 465) for the bibliometric analysis. For each unique study, we extracted the addresses of all authors. For corresponding authors, we used the address provided through Web of Science or extracted from the paper manually. Some publications in our dataset did not have an identifiable corresponding author, in which case it was left empty. In cases where the corresponding author had multiple affiliations, we considered only the first. For each study, we identified the unique affiliation addresses and the number of authors affiliated to each address. We ran all unique addresses from papers through a Google Maps-based geocoding software on Google Sheets, Awesome Table Geocode, that returns latitude and longitude for each address (https://awesome-table.com).

To analyse collaboration potential within the region for future syntheses, we compared the geographic positions of the authors and the papers in different ways. For these analyses, we limited the dataset to only those studies where we had information on the location of the authors as well as the site (n=456). First, we examined the proportion of authors in each study that were affiliated with institutions in India and its relation to the area sampled. Second, we examined the distance of the study site from the corresponding author institution for each study, by calculating the Haversine distance (the shortest distance between the two points assuming a spherical earth) between the two pairs of coordinates using the *geosphere* package (Hijmans, 2022a). We then looked at how this distance has changed over the years.

All analyses were performed in R version 4.1.3 (R Core Team, 2022) using packages *dplyr* and *tidyverse* (Wickham et al., 2019, 2022).

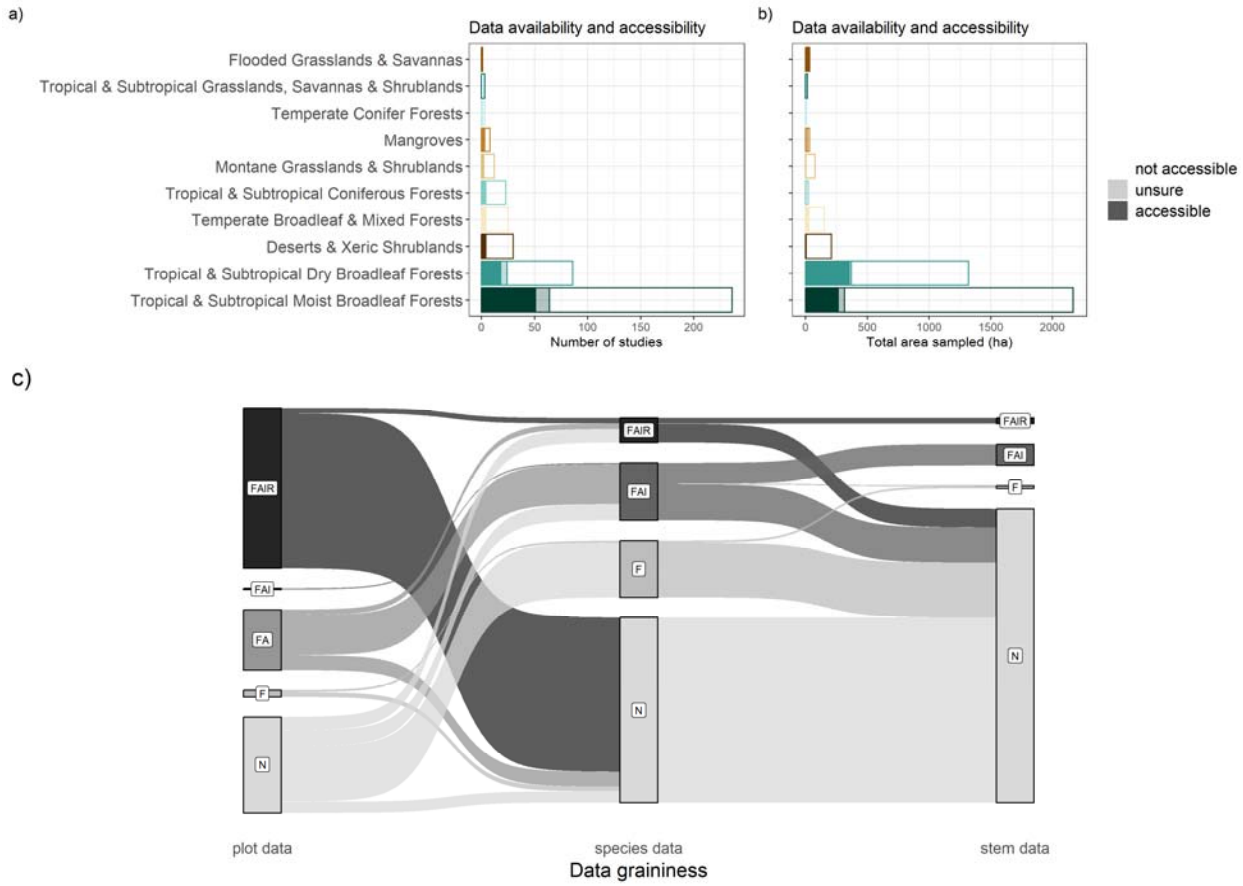

Figure 2: Data availability and accessibility across biomes. *a) Number of studies and b)total area sampled across the 10 vegetated biomes of India among tree inventories studies in India. Solid bars represent data accessible in repositories or in the appendix while empty bars represent inaccessible data. Lightly coloured bars represent studies where the accessibility status was uncertain. Studies focussing on a single biome are only reported here. See supplementary material for details. c) Data distribution along data graininess and Findable, Accessible, Interoperable, Reusable (FAIR) principles of data reuse. Data reported only for studies where data was accessible for at least one of the categories of data graininess, the rest has been excluded from this figure. Legends F, FA, FAI, FAIR represent position along the reusability scale as defined by these principles. N represents data that were not findable. Size of the bars represent total sampled area in each of these categories. For detailed methods, refer to the supplementary material.*

RESULTS

DATA AVAILABILITY AND ACCESSIBILITY

The 465 studies reported in the journal articles in the INvenTree dataset span 10 biomes and cover 31 of the 36 states and union territories of India (Fig 1). We identified 436 studies of this

dataset that report on area-based inventories, e.g. transects or plot-based measures, sampling a total of 4653.64 ha (Fig 1). The remaining were checklists, duplicates or did not report key metrics to estimate sampling effort. The most commonly studied biome was Tropical & Subtropical Moist Broadleaf Forests (236 out of 465, 50.75%), followed by Tropical & Subtropical Dry Broadleaf Forests (86 out of 465, 18.49%) (Fig 2a). In terms of area, the most extensively sampled biome was Tropical & Subtropical Moist Broadleaf Forests with a total area of 2169.3 ha sampled across 221 studies (Fig 2b).

Most studies in the INvenTree dataset had low effort in terms of total area sampled; the median sampled area of the studies was 2 ha. Of these, 84.73% report on datasets that sample less than a total area of 100 ha, 72.47% less than 10 ha, 27.96% less than 1 ha. Across these studies, data accessibility was poor; only 101 studies, representing 21.72% of the studies and 14.8% of the area had data that was available either as a table, appendix, supplementary information or archived in a repository. Sampled area was not statistically significantly different between data-accessible and data-inaccessible studies (one-way ANOVA, df=1, SS=0.22, p=0.45). Data access was fairly evenly limited across the bioregions studied (-squared test statistic for number of studies = 40, p=0.1, for area = 80, p=0.2) (Fig 2b).

While examining the reusability of the datasets under the FAIR framework, we found considerable variation in the level of detail of findable and reusable data across studies that had "accessible" data (Fig 2c). Much of the datasets classified broadly as "accessible" (14.8% of the total sampled area) met reusability standards only for plot-level (12.81% of total sampled area across studies) or species-level data (1.95% of sampled area). Among the studies assessed, only 5 representing 0.47% of sampled area met reusability standards for stem level data, the finest grain of tree inventory information assessed. Despite limited data accessibility, the text of 56.77% of published articles were openly accessible (Fig S4). Studies were most commonly published in the journal Tropical Ecology (n=40) (Fig S4).

AUTHORSHIP PATTERNS

Over 4 in 5 of the corresponding authors of papers in the INvenTree dataset were based in institutions in India (82.8%), with a small proportion of corresponding authors (14.19%) from other countries, primarily the United States (4.73%), France (2.37%) and China (1.72%) (Fig 3). Within India, affiliations were spread widely across the country, showing authorship and data ownership across institutions (Fig 3a). The number of publications have grown over the years and much of this increase is attributable to corresponding authors based in India (Fig S13). A clear majority, 1612 out of a total 1870 authors across articles in the meta-dataset were affiliated with Indian institutions and organisations ('domestic' in Fig 3), but 24.95% of papers also had authors affiliated with non-Indian institutions ('foreign' in Fig 3). When scored for the percentage of Indian-affiliated authors in a study, 90.11% of studies had more than 50% authors affiliated to India. 73.33% studies had all authors affiliated to Indian institutions. Although

large-scale studies were few, they tended to have high or near-total in-country authorship (Fig 3). We observed an increasing trend of large-scale studies over the years (S13), but local and small-scale studies also continue to be published (Fig 3). Further, most studies in the INvenTree dataset were conducted locally, close to the institution of the corresponding author; the median distance between corresponding author affiliation and study site was 280.38 km (Fig 3).

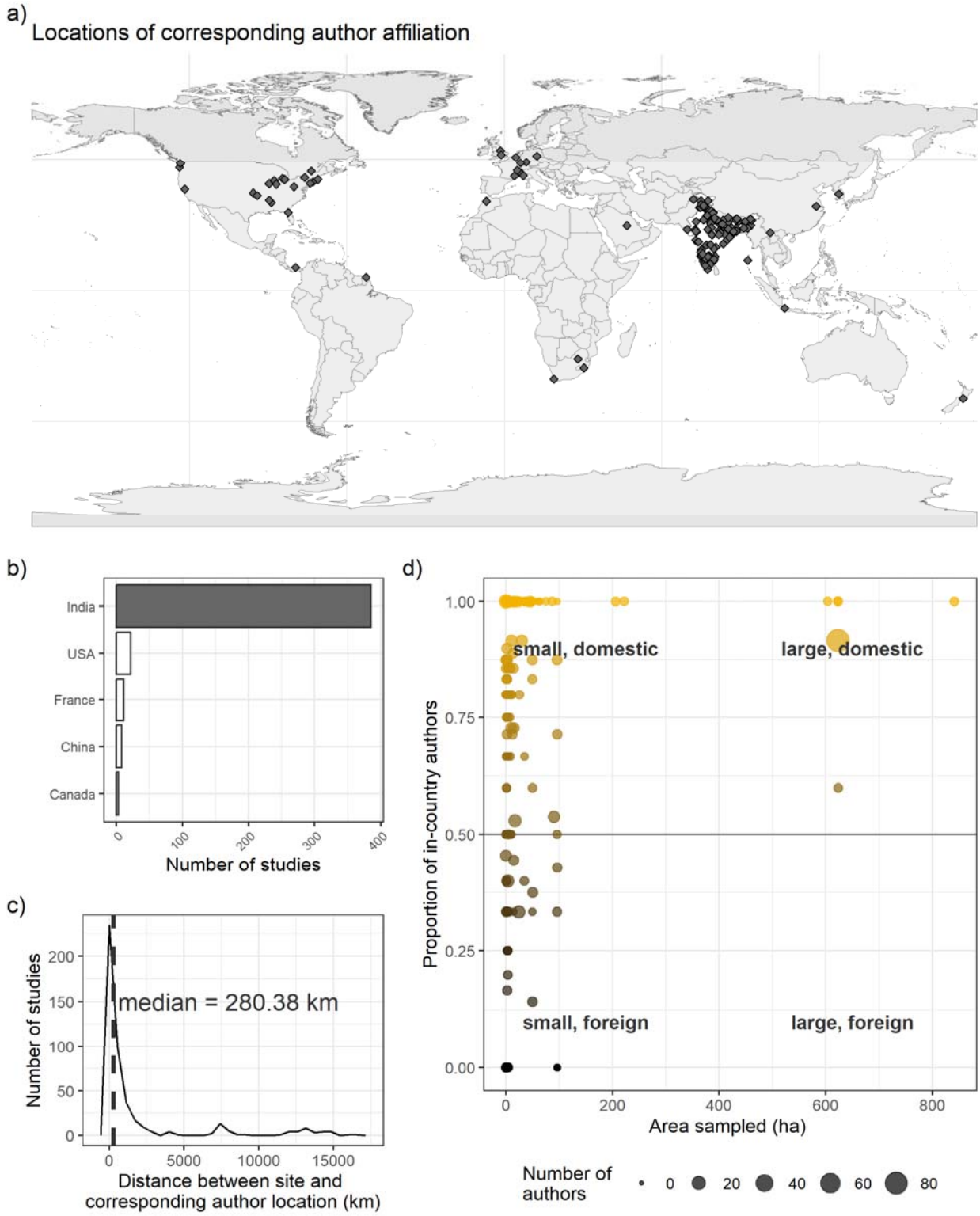
a)
Locations of corresponding author affiliation
b)
India
USA
France
China
Canada
0
100
200
300
400
Number of studies
c)
Number of studies
200
150
100
50
0
median = 280.38 km
0
5000
10000
15000
Distance between site and corresponding author location (km)
d)
Proportion of in-country authors
1.00
0.75
0.50
0.25
0.00
small, domestic
large, domestic
small, foreign
large, foreign
0
200
400
600
800
Area sampled (ha)
Number of authors
0
20
40
60
80

Figure 3: INvenTree authorship patterns and data equity. *a) Locations of corresponding authors of papers in the INvenTree dataset, b) top five countries, representing 92.26%, where corresponding authors were affiliated; c) Measuring equity within the country - distance between corresponding author affiliation and study site across studies. d) Investigating parachute research - proportion of domestic and foreign authors in the INvenTree dataset and their relationship to area sampled. Size of the points represents number of authors in the study, colour is indicative of the proportion of in-country authors.*

Sampling priorities

The INvenTree dataset was used to identify sampling priority regions for ecological and conservation research (Fig 4). The INvenTree dataset showed disproportionate sampling across districts, with most (223 out of 341) districts with tree cover never sampled (Fig S6). When sampling gaps were visualised along tree cover and tree cover loss gradients, high heterogeneity among sites and biomes emerged (Fig 4). Notably, among districts with high tree cover, the northern Western Ghats, several districts of North-east India, parts of the Western Himalayas, and a few districts in eastern India fell in the lowest quantile of area sampled (Fig 4, Fig S6). Ecological sampling priority regions that fall in lower tree cover quantiles were many of the drier districts in peninsular India like the Deccan plateau, Eastern Ghats, Central India, as well as the western Himalayan foothills and the Nicobar Islands. There were also a few datasets reporting on tree-based inventories that fell outside of these forested districts (Fig S5). When considering sampling effort across ecoregions, the effort was highly skewed and most regions were highly undersampled relative to their extent (Fig S12).

Districts of the highest conservation sampling priority were largely different from those of the ecological sampling priority districts. Districts in the highest quantile of forest loss relative to forest cover were clustered mainly in north-east India and the eastern states of Odisha, Andhra Pradesh, Telengana and Chattisgarh (Fig S9). Among districts that fell in the highest quantile of tree cover loss, districts in the north-east and eastern India were the least studied (Fig 4). Data availability from areas of tree cover loss was poor from the northern Western Ghats, northeast India and the Nicobar Islands.

Discussion

The INvenTree dataset, to our knowledge, is the first and largest meta-dataset of tree inventory information from India. This dataset covers all tree-based biomes of India and spans a total of 4653.64 ha. Using this dataset, we identify geographic and biome gaps in sampling and define ecological and conservation sampling priority regions. Most studies in the INvenTree dataset are authored by researchers affiliated to Indian universities and study local ecosystems. Although 80% of the dataset is not accessible, the strong presence of researchers in Indian organisations who are owners of high quality local data indicates potential for a regional network to promote a better understanding of ecological processes across larger spatial scales.

Harmonising these datasets under fair sharing protocols is a first step towards synthetic research in the region. Given India's high biodiversity and human-dominated landscapes, such a network could promote a better understanding of ecosystems in the Anthropocene, while benefiting research and researchers.

Using the INvenTree dataset, we identified the extent and sampling biases across biomes and regions, providing directions to align future sampling of tree-based biomes within the country (Fig 4). The INvenTree dataset spans all of India's diverse tree-based biomes, but shows non-uniform sampling across these. Among biomes, tropical and subtropical moist and dry forests were well sampled relative to their area. However, within India, there has been increased attention on the diversity of tropical dry forest types (Ratnam et al., 2016), their high biodiversity and threats to their conservation (Nerlekar et al., 2022, 2024). Our result highlights potential for these tree inventory data, once collated and standardized, to contribute to our larger understanding of dry tropical forests which remain among the least understood and most threatened tropical tree biomes globally (Miles et al., 2006; Siyum, 2020). Although the moist forests were better sampled, the distribution of the plots was clustered in the southern western ghats and north eastern forests, which are more speciose (Fig 1). Consequently, we report ecological sampling priority in the northern Western Ghats and north-east India and conservation sampling priority in north-east India as well as eastern India (Fig 4). Regional disparities in sampling can lead to biased understanding of ecological heterogeneity in patterns and processes which may result in inappropriate management, conservation or restoration practices. It is concerning that our proposed regions of priority conservation sampling (high loss/low sampling) overlap regions of large conservation concern with megaprojects (e.g. dams and oil palm expansion in North-East India, Bhattacharya (2022), Srinivasan et al. (2021); proposed development in the Nicobar Islands, Ramachandran (2022); mineral resource extraction in eastern India, Aggarwal (2021)). As these regions continue to lose tree cover, it is an urgent need to collect baseline information to support stakeholders in appropriate conservation action. Our proposed ecological and conservation sampling priority maps can provide a blueprint for the next decades of tree sampling in the country to help refine our ecological understanding and inform management and conservation.

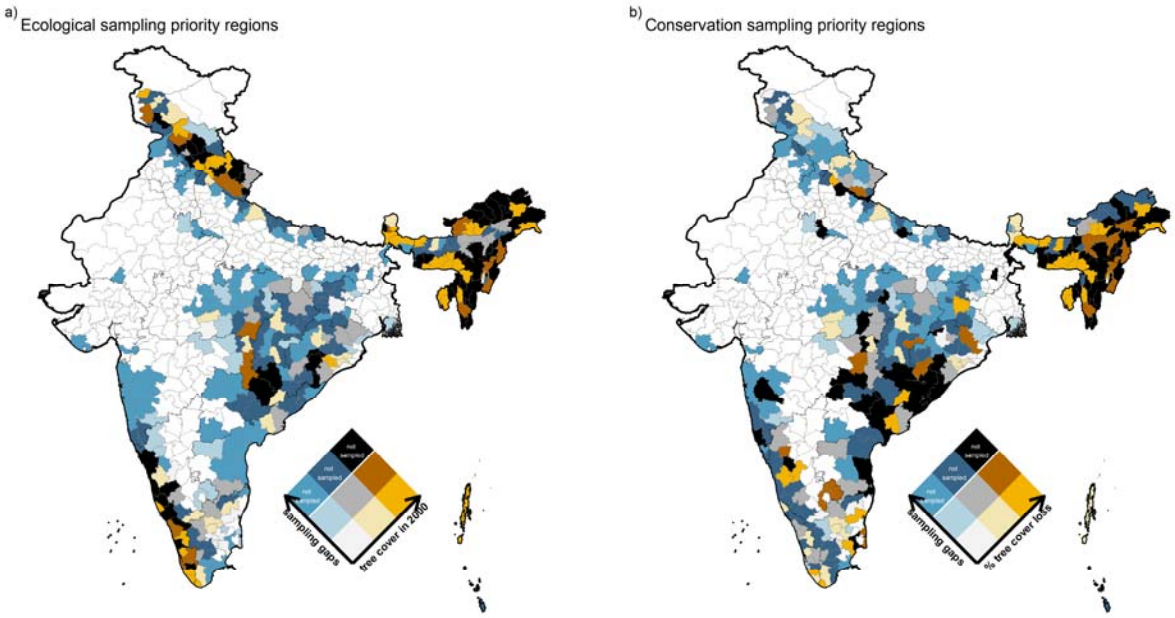

Figure 4: Ecological and conservation sampling priority regions. *Bivariate chloropleth maps showing sampling gaps in plot-based tree inventories across districts in India in relation to a) mean tree cover in the district in 2000 and b) mean tree cover loss from 2000-2021 averaged over the area of the district. Sampling gaps were categorised into three categories; districts with 0 sampling effort were coded as the highest gap, the rest were split into two groups. Refer methods and supplementary for details.*

Table 1 | Potential applications emerging from a combined, harmonised INvenTree

Category	Applications	Examples
Basic ecological research	Understanding patterns and drivers of taxonomic and phylogenetic diversity	Page and Shanker (2020)
	Patterns in ecosystem function (productivity, temporal and spatial stability, carbon balance, etc.) across heterogenous landscapes	Osuri et al. (2020)
	Ecological forecasting of ecosystem functions with climate	
Conservation and policy	Monitoring, evaluation and learning on forest restoration and other nature based solutions for climate change mitigation	REDD+, carbon credits
	Habitat assessment for in situ conservation of keystone species	Koulgi et al. (2019)
	Aiding natural resource management and measuring its effectiveness (based on different metrics such as carbon stocks, biodiversity, etc)	

Table 1 | Potential applications emerging from a combined, harmonised INvenTree

Category	Applications	Examples
	Policies to promote future research in underserved regions	Critical Ecosystem Partnership Fund
Socio-cultural dimensions	Enable sustainable use and biocultural conservation of important forest resources	Pradhan and Ormsby (2020)
	Eco-tourism and other recreational uses	
	Potential for data-based teaching modules on tropical ecology	Organisation for Tropical Studies Field Courses

CHALLENGES AND PROMISES OF SYNTHESISING INVENTREE DATA

A harmonised database of tree inventories from India, beyond the INvenTree meta-dataset, can fill crucial knowledge gaps in the human dimension of global tropical ecology to aid research, conservation and socio-cultural dimensions (Table 1). Large-scale datasets are more than the sum of their parts; for example, data published by Ramesh et al. (2010), covering the Western Ghats, has allowed for several syntheses, both at regional and global scales ranging from patterns and drivers of diversity to drivers of above ground carbon storage(e.g. Hardy et al. (2012); Krishnadas et al. (2021); Osuri et al. (2020); Davidar et al. (2018)). Spatially distributed datasets like the Forest Inventory and Analysis programme in the United States have led to at least 180 publications and increased the accuracy of predictive modeling of ecological communities (Tinkham et al., 2018). Similar to other national programmes, The Forest Survey of India (FSI) conducts country-wide tree inventories spanning ~ 3500 ha total area every 5 years with a systematic gridded approach (https://fsi.nic.in/about-forest-inventory, Tewari (2016)). In comparison, studies in the INvenTree dataset in aggregate covers more than this area even after accounting for sites sampled multiple times. A harmonised database such as INvenTree will enable synthetic research on the patterns and processes structuring ecosystems across the region. Moreover, datasets with repeat sampling could also lend to ecological forecasting of ecosystems states with climate (e.g. Heilman et al. (2022)). Such data could immensely improve conservation efforts both by government and private actors by improving data available for monitoring and evaluation of forest management, restoration efforts, carbon credit projects and other nature-based solutions to mitigate climate change. These data would also be useful for universities and educators, making it possible to create data-driven educational products for ecology courses in India (e.g. Organisation for Tropical Studies courses in other geographies). We also envision that INvenTree data would serve as crucial baselines for practitioners and local communities involved in sustainable use of forest resources as well as "biocultural"

conservation that balances conservation goals with local cultural values and worldviews (Sterling et al., 2017).

The INvenTree dataset highlights the potential for collaborative approaches within the country to facilitate inclusion of high quality data into regional and global syntheses. Most research represented in the INvenTree dataset was conducted locally, likely integrating deep knowledge of landscape context, taxonomy and natural history. In heterogeneous tropical landscapes like India with high alpha and beta diversity, such localized, specialized data collection can be crucial for data quality. However, at present, the reusability of these data for further analyses is hindered by the small scale of individual datasets and their distributed ownership. Since many of these datasets have total sampling effort < 2 ha, the effort required to individually request permissions from authors and harmonise these may outweigh the benefits of their inclusion into previous global syntheses. An accepted set of best practices for data reusability is described by the FAIR principles for data; Findable, Accessible, Interoperable and Reusable (Wilkinson et al., 2016). Our assessment of the reusability of these datasets revealed that even while datasets were "findable" and "accessible", the current forms in which they are publicly available most often did not meet reusability standards, especially for the finest grain, stem-level data, required for many syntheses. Despite the limited access at the stem level, plot-based vegetation studies are multi-layered in their data as well as further utility, allowing useful, scalable, synthetic research even with plot and species level data. Harmonising these heterogeneous datasets, to achieve "Interoperability" and "Reusability", is difficult as they can be conducted using slightly different methodologies or adapted to context. Moreover, standardising species names across studies, including dealing with local names, requires taxonomic expertise and biogeographic knowledge (Singh, 2008). Assembling such a database would require close, continuous, collaborations with data owners and collectors.

The distributed inventories collated in INvenTree are important not just because of their extent, but the perspective of data sovereignty, and equity. FAIR principles for data sharing are not always fair for tropical forest data owners and making data unconditionally open could lead to further extractive science, as quantitative skills and publishing power tends to be concentrated in the global north and in elite universities (De Lima et al., 2022; Mori et al., 2015). Existing networks such as ForestPlots, respond to this inequity by insisting on co-authorship for primary data collectors (ForestPlots.net et al., 2021). We are encouraged by our findings that showcase a steady pattern of in-country authorship on tree-inventory-based publications from India, a positive contrast to the global pattern of parachute research in tropical landscapes (Asase et al., 2022; Miller et al., 2023). Acknowledging the enormous efforts and potential needs of current and future data owners, we envision a regional network that prioritises fair sharing policies as well as quantitative processing tools (e.g. open source code, tutorials, periodic data workshops). Combining frameworks for data sharing with tools and networks can empower Indian researchers, especially early-career researchers and those in under-funded universities to access and utilise these data for their research.

In conclusion, using the first synthesized metadataset of tree inventory data from India, we identified geographic gaps in sampling and highlighted priority regions for future sampling to fill critical gaps in tropical ecology. We also identify strong scholarship in forest ecology within India, demonstrated by a large number of authors and data owners affiliated to Indian institutions. Based on our findings, we propose The India Tree Inventory Network, a regional network of forest researchers from India that will promote synthetic forest research in India while prioritising data sovereignty and knowledge equity at the global and national scale. The INvenTree network will build on existing frameworks for data and knowledge sharing that acknowledge diverse contributions, costs and benefits inherent to an unequal world (Cooke et al., 2021; De Lima et al., 2022; Ocampo-Ariza et al., 2023).

POSITIONALITY STATEMENT

All authors of this paper are Indian nationals, working in ecology and allied fields, with substantial field experience in India. KA and NM are currently postdoctoral researchers in US institutions and were also affiliated with US universities for their PhD, but have studied and worked in Indian institutions up to the Masters level. KA has led field data collection in India through all stages of her professional life starting from undergrad. AG and AK are PhD students in Indian institutions. AS is affiliated with US institutions for his PhD and has studied and worked in India until the Masters level. AJ is a Masters student in Europe. TN is a PhD student in the UK doing field work in India and has studied in India until her undergraduate degree. SO is a PhD student in Germany. MS is a Professor at an Indian institution.

ACKNOWLEDGEMENTS

We are grateful to the community of data collectors and owners contributing to forest ecology in India. KA acknowledges the support of the Smithsonian Institution Postdoctoral Fellowship during the completion of this manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA ACCESSIBILITY STATEMENT

The data and code used in this manuscript will be made available on the India-INvenTree GitHub repository (www.github.com/India-INvenTree/metadata-analysis) and on Zenodo upon acceptance of this manuscript.

References

Aggarwal, M. (2021). Environmental degradation in India's oldest coal mining belt leaves locals gasping for relief. In *Mongabay-India*. https://india.mongabay.com/2021/03/environmental-degradation-in-indias-oldest-coal-mining-belt-leaves-locals-gasping-for-relief/.

Amano, T., & Sutherland, W. J. (2013). Four barriers to the global understanding of biodiversity conservation: Wealth, language, geographical location and security. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1756), 20122649. https://doi.org/10.1098/rspb.2012.2649

Anderson-Teixeira, K. J., Davies, S. J., Bennett, A. C., Gonzalez-Akre, E. B., Muller-Landau, H. C., Joseph Wright, S., Abu Salim, K., Almeyda Zambrano, A. M., Alonso, A., Baltzer, J. L., Basset, Y., Bourg, N. A., Broadbent, E. N., Brockelman, W. Y., Bunyavejchewin, S., Burslem, D. F. R. P., Butt, N., Cao, M., Cardenas, D., … Zimmerman, J. (2015). CTFS-ForestGEO: A worldwide network monitoring forests in an era of global change. *Global Change Biology*, *21*(2), 528–549. https://doi.org/10.1111/gcb.12712

Anderson-Teixeira, K. J., Wang, M. M. H., McGarvey, J. C., Herrmann, V., Tepley, A. J., Bond-Lamberty, B., & LeBauer, D. S. (2018). ForC : A global database of forest carbon stocks and fluxes. *Ecology*, *99*(6), 1507–1507. https://doi.org/10.1002/ecy.2229

Asase, A., Mzumara-Gawa, T. I., Owino, J. O., Peterson, A. T., & Saupe, E. (2022). Replacing "parachute science" with "global science" in ecology and conservation biology. *Conservation Science and Practice*, *4*(5), e517. https://doi.org/10.1111/csp2.517

Barlow, J., Gardner, T. A., Lees, A. C., Parry, L., & Peres, C. A. (2012). How pristine are tropical forests? An ecological perspective on the pre-Columbian human footprint in Amazonia and implications for contemporary conservation. *Biological Conservation*, *151*(1), 45–49. https://doi.org/10.1016/j.biocon.2011.10.013

Bhattacharya, S. (2022). *India Wants Its Northeast To Be Its "Powerhouse." What of the Indigenous People? – The Wire Science*.

Briggs, J. C. (2003). The biogeographic and tectonic history of India. *Journal of Biogeography*, *30*(3), 381–388. https://doi.org/10.1046/j.1365-2699.2003.00809.x

Carvalho, R. L., Resende, A. F., Barlow, J., França, F. M., Moura, M. R., Maciel, R., Alves-Martins, F., Shutt, J., Nunes, C. A., Elias, F., Silveira, J. M., Stegmann, L., Baccaro, F. B., Juen, L., Schietti, J., Aragão, L., Berenguer, E., Castello, L., Costa, F. R. C., … Feitosa, Y. (2023). Pervasive gaps in Amazonian ecological research. *Current Biology*, *33*(16), 3495–3504.e4. https://doi.org/10.1016/j.cub.2023.06.077

Cazzolla Gatti, R., Reich, P. B., Gamarra, J. G. P., Crowther, T., Hui, C., Morera, A., Bastin, J.-F., de-Miguel, S., Nabuurs, G.-J., Svenning, J.-C., Serra-Diaz, J. M., Merow, C., Enquist, B., Kamenetsky, M., Lee, J., Zhu, J., Fang, J., Jacobs, D. F., Pijanowski, B., … Liang, J. (2022). The number of tree species on Earth. *Proceedings of the National Academy of Sciences*, *119*(6), e2115329119. https://doi.org/10.1073/pnas.2115329119

Champion, S. H. G., & Seth, S. K. (1968). *A Revised Survey of the Forest Types of India*. Delhi: Government of India Press.

Cooke, S. J., Nguyen, V. M., Young, N., Reid, A. J., Roche, D. G., Bennett, N. J., Rytwinski, T., & Bennett, J. R. (2021). Contemporary authorship guidelines fail to recognize diverse contributions in conservation science research. *Ecological Solutions and Evidence*, *2*(2), e12060. https://doi.org/10.1002/2688-8319.12060

Crowther, T. W., Glick, H. B., Covey, K. R., Bettigole, C., Maynard, D. S., Thomas, S. M., Smith, J. R., Hintler, G., Duguid, M. C., Amatulli, G., Tuanmu, M.-N., Jetz, W., Salas, C., Stam, C., Piotto, D., Tavani, R., Green, S., Bruce, G., Williams, S. J., … Bradford, M. A. (2015). Mapping tree density at a global scale. *Nature*, *525*(7568), 201–205. https://doi.org/10.1038/nature14967

Davidar, P., Munoz, F., Puyravaud, J.-P., Mohandass, D., & Ramachandran, V. S. (2018). Multiple facets of rarity among rain forest trees in the Western Ghats of India. *Biological Conservation*, *228*, 110–119. https://doi.org/10.1016/j.biocon.2018.10.014

De Lima, R. A. F., Phillips, O. L., Duque, A., Tello, J. S., Davies, S. J., De Oliveira, A. A., Muller, S., Honorio Coronado, E. N., Vilanova, E., Cuni-Sanchez, A., Baker, T. R., Ryan, C. M., Malizia, A., Lewis, S. L., Ter Steege, H., Ferreira, J., Marimon, B. S., Luu, H. T., Imani, G., … Vásquez, R. (2022). Making forest data fair and open. *Nature Ecology & Evolution*, *6*(6), 656–658. https://doi.org/10.1038/s41559-022-01738-7

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., … Saleem, M. (2017). An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience*, *67*(6), 534–545. https://doi.org/10.1093/biosci/bix014

Farley, S. S., Dawson, A., Goring, S. J., & Williams, J. W. (2018). Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions. *BioScience*, *68*(8), 563–576. https://doi.org/10.1093/biosci/biy068

ForestPlots.net, Blundo, C., Carilla, J., Grau, R., Malizia, A., Malizia, L., Osinaga-Acosta, O., Bird, M., Bradford, M., Catchpole, D., Ford, A., Graham, A., Hilbert, D., Kemp, J., Laurance, S.,

Laurance, W., Ishida, F. Y., Marshall, A., Waite, C., … Tran, H. D. (2021). Taking the pulse of Earth's tropical forests using networks of highly distributed plots. *Biological Conservation*, *260*, 108849. https://doi.org/10.1016/j.biocon.2020.108849

Fraser, J. A., Cosiaux, A., Walters, G., Asiyanbi, A., Osei-Wusu Adjei, P., Addo-Fordjour, P., Fairhead, J., Kialo, P., Laurier Engone Obiang, N., & Oslisly, R. (2024). Defining the Anthropocene tropical forest: Moving beyond "disturbance" and "landscape domestication" with concepts from African worldviews. *The Anthropocene Review*, 20530196231226307. https://doi.org/10.1177/20530196231226307

Friedl, M., & Sulla-Menashe, D. (2022). *MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061*. [object Object]. https://doi.org/10.5067/MODIS/MCD12Q1.061

Gadgil, M., & Thapar, R. (1990). Human Ecology in India Some Historical Perspectives. *Interdisciplinary Science Reviews*, *15*(3), 209–223. https://doi.org/10.1179/isr.1990.15.3.209

Hagen, O., Skeels, A., Onstein, R. E., Jetz, W., & Pellissier, L. (2021). Earth history events shaped the evolution of uneven biodiversity across tropical moist forests. *Proceedings of the National Academy of Sciences*, *118*(40), e2026347118. https://doi.org/10.1073/pnas.2026347118

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Hardy, O. J., Couteron, P., Munoz, F., Ramesh, B. R., & Pélissier, R. (2012). Phylogenetic turnover in tropical tree communities: Impact of environmental filtering, biogeography and mesoclimatic niche conservatism. *Global Ecology and Biogeography*, *21*(10), 1007–1016. https://doi.org/10.1111/j.1466-8238.2011.00742.x

Heilman, K. A., Dietze, M. C., Arizpe, A. A., Aragon, J., Gray, A., Shaw, J. D., Finley, A. O., Klesse, S., DeRose, R. J., & Evans, M. E. K. (2022). Ecological forecasting of tree growth: Regional fusion of tree-ring and forest inventory data to quantify drivers and characterize uncertainty. *Global Change Biology*, *28*(7), 2442–2460. https://doi.org/10.1111/gcb.16038

Hijmans, R. J. (2022a). *Geosphere: Spherical trigonometry* [Manual].

Hijmans, R. J. (2022b). *Terra: Spatial data analysis* [Manual].

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, *44*(9), 1259–1269. https://doi.org/10.1111/ecog.05926

Jung, M., Arnell, A., De Lamo, X., García-Rangel, S., Lewis, M., Mark, J., Merow, C., Miles, L., Ondo, I., Pironon, S., Ravilious, C., Rivers, M., Schepaschenko, D., Tallowin, O., Van Soesbergen, A., Govaerts, R., Boyle, B. L., Enquist, B. J., Feng, X., … Visconti, P. (2021). Areas of global importance for conserving terrestrial biodiversity, carbon and water. *Nature Ecology & Evolution*, *5*(11), 1499–1509. https://doi.org/10.1038/s41559-021-01528-7

Karanth, K. K., & DeFries, R. (2010). Conservation and management in human-dominated landscapes: Case studies from India. *Biological Conservation*, *143*(12), 2865–2869. https://doi.org/10.1016/j.biocon.2010.05.002

Koulgi, P. S., Clinton, N., & Karanth, K. K. (2019). Extensive vegetation browning and drying in forests of India's Tiger Reserves. *Scientific Reports*, *9*(1), 14976. https://doi.org/10.1038/s41598-019-51118-8

Krishnadas, M., Sankaran, M., Page, N., Joshi, J., Machado, S., Nataraj, N., Chengappa, S. K., Kumar, V., Kumar, A., & Krishnamani, R. (2021). Seasonal drought regulates species distributions and assembly of tree communities across a tropical wet forest region. *Global Ecology and Biogeography*, *30*(9), 1847–1862. https://doi.org/10.1111/geb.13350

Kwon, D. (2022). Open-access publishing fees deter researchers in the global south. *Nature*, d41586-022-00342-w. https://doi.org/10.1038/d41586-022-00342-w

Maas, B., Pakeman, R. J., Godet, L., Smith, L., Devictor, V., & Primack, R. (2021). Women and Global South strikingly underrepresented among top-publishing ecologists. *Conservation Letters*, *14*(4), e12797. https://doi.org/10.1111/conl.12797

Malhi, Y., Phillips, O. L., Lloyd, J., Baker, T., Wright, J., Almeida, S., Arroyo, L., Frederiksen, T., Grace, J., Higuchi, N., Killeen, T., Laurance, W. F., Leaño, C., Lewis, S., Meir, P., Monteagudo, A., Neill, D., Núñez Vargas, P., Panfil, S. N., … Vinceti, B. (2002). An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR). *Journal of Vegetation Science*, *13*(3), 439–450. https://doi.org/10.1111/j.1654-1103.2002.tb02068.x

Mani, M. S. (1974). *Ecology and Biogeography in India* (Vol. 12). The Hague: Dr W. Junk b.v. Publishers.

Martin, L. J., Blossey, B., & Ellis, E. (2012). Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, *10*(4), 195–201. https://doi.org/10.1890/110154

Mekonnen, A., Downs, C., Effiom, E. O., Kibaja, M., Lawes, M. J., Omeja, P., Ratsoavina, F. M., Razafindratsima, O., Sarkar, D., Stenseth, N. C., & Chapman, C. A. (2021). Can I afford to publish? A dilemma for African scholars. *Ecology Letters*. https://doi.org/10.1111/ele.13949

Miles, L., Newton, A. C., DeFries, R. S., Ravilious, C., May, I., Blyth, S., Kapos, V., & Gordon, J. E. (2006). A global overview of the conservation status of tropical dry forests. *Journal of Biogeography*, *33*(3), 491–505. https://doi.org/10.1111/j.1365-2699.2005.01424.x

Miller, J., White, T. B., & Christie, A. P. (2023). Parachute conservation: Investigating trends in international research. *Conservation Letters*, *16*(3), e12947. https://doi.org/10.1111/conl.12947

Mittermeier, R. A., Mittermeier, R. A., & Cemex, S.A. de C.V (Eds.). (2004). *Hotspots revisited* (First English Edition). Cemex.

Mori, A. S., Qian, S., & Tatsumi, S. (2015). Academic inequality through the lens of community ecology: A meta-analysis. *PeerJ*, *3*, e1457. https://doi.org/10.7717/peerj.1457

Mugal, M. A., Wani, S. A., Dar, F. A., Islam, T., Gulzar, R., Malik, A. H., Reddy, C. S., & Khuroo, A. A. (2023). Bridging global knowledge gaps in biodiversity databases: A comprehensive data synthesis on tree diversity of India. *Biodiversity and Conservation*, *32*(10), 3089–3107. https://doi.org/10.1007/s10531-023-02659-y

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Nerlekar, A. N., Chorghe, A. R., Dalavi, J. V., Kusom, R. K., Karuppusamy, S., Kamath, V., Pokar, R., Rengaian, G., Sardesai, M. M., & Kambale, S. S. (2022). Exponential rise in the discovery of endemic plants underscores the need to conserve the Indian savannas. *Biotropica*, *54*(2), 405–417. https://doi.org/10.1111/btp.13062

Nerlekar, A. N., Munje, A., Mhaisalkar, P., Hiremath, A. J., & Veldman, J. W. (2024). Tillage agriculture and afforestation threaten tropical savanna plant communities across a broad rainfall gradient in India. *Journal of Ecology*, *112*(1), 98–109. https://doi.org/10.1111/1365-2745.14221

Nuñez, M. A., Chiuffo, M. C., Pauchard, A., & Zenni, R. D. (2021). Making ecology really global. *Trends in Ecology & Evolution*, *36*(9), 766–769. https://doi.org/10.1016/j.tree.2021.06.004

Ocampo-Ariza, C., Toledo-Hernández, M., Librán-Embid, F., Armenteras, D., Vansynghel, J., Raveloaritiana, E., Arimond, I., Angulo-Rubiano, A., Tscharntke, T., Ramírez-Castañeda, V., Wurz, A., Marcacci, G., Anders, M., Urbina-Cardona, J. N., De Vos, A., Devy, S., Westphal, C., Toomey, A., Sheherazade, … Maas, B. (2023). Global South leadership towards inclusive tropical ecology and conservation. *Perspectives in Ecology and Conservation*, *21*(1), 17–24. https://doi.org/10.1016/j.pecon.2023.01.002

Odeny, B., & Bosurgi, R. (2022). Time to end parachute science. *PLOS Medicine*, *19*(9), e1004099. https://doi.org/10.1371/journal.pmed.1004099

Osuri, A. M., Machado, S., Ratnam, J., Sankaran, M., Ayyappan, N., Muthuramkumar, S., Parthasarathy, N., Pélissier, R., Ramesh, B. R., DeFries, R., & Naeem, S. (2020). Tree diversity and carbon storage cobenefits in tropical human-dominated landscapes. *Conservation Letters*, *13*(2), e12699. https://doi.org/10.1111/conl.12699

Page, N. V., & Shanker, K. (2020). Climatic stability drives latitudinal trends in range size and richness of woody plants in the Western Ghats, India. *PLOS ONE*, *15*(7), e0235733. https://doi.org/10.1371/journal.pone.0235733

Poulter, B., Aragão, L., Andela, N., Bellassen, V., Ciais, P., Kato, T., Lin, X., Nachin, B., Luyssaert, S., Pederson, N., Peylin, P., Piao, S., Pugh, T., Saatchi, S., Schepaschenko, D., Schelhaas, M., & Shivdenko, A. (2019). *The global forest age dataset and its uncertainties (GFADv1.1)* [Data Set]. PANGAEA / NASA National Aeronautics and Space Administration. https://doi.org/10.1594/PANGAEA.897392

Pradhan, A., & Ormsby, A. A. (2020). Biocultural conservation in the sacred forests of Odisha, India. *Environmental Conservation*, *47*(3), 190–196. https://doi.org/10.1017/S0376892920000181

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Raja, N. B., Dunne, E. M., Matiwane, A., Khan, T. M., Nätscher, P. S., Ghilardi, A. M., & Chattopadhyay, D. (2021). Colonial history and global economics distort our understanding of deep-time biodiversity. *Nature Ecology & Evolution*, *6*(2), 145–154. https://doi.org/10.1038/s41559-021-01608-8

Ramachandran, T. (2022). [Maps] Environmental path cleared for Great Nicobar mega project. In *Mongabay-India*. https://india.mongabay.com/2022/10/maps-environmental-path-cleared-for-great-nicobar-mega-project/.

Ramesh, B. R., Swaminath, M. H., Patil, S. V., Dasappa, Pélissier, R., Venugopal, P. D., Aravajy, S., Elouard, C., & Ramalingam, S. (2010). Forest stand structure and composition in 96 sites along environmental gradients in the central Western Ghats of India: Ecological Archives E091-216. *Ecology*, *91*(10), 3118–3118. https://doi.org/10.1890/10-0133.1

Ramírez-Castañeda, V. (2020). Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science: The case of Colombian researchers in biological sciences. *PLoS One*, *15*(9), e0238372. https://doi.org/http://dx.doi.org.ezproxy.cul.columbia.edu/10.1371/journal.pone.0238372

Ratnam, J., Tomlinson, K. W., Rasquinha, D. N., & Sankaran, M. (2016). Savannahs of Asia: Antiquity, biogeography, and an uncertain future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1703), 20150305. https://doi.org/10.1098/rstb.2015.0305

Reddy, C. S., Jha, C. S., Diwakar, P. G., & Dadhwal, V. K. (2015). Nationwide classification of forest types of India using remote sensing and GIS. *Environmental Monitoring and Assessment*, *187*(12), 777. https://doi.org/10.1007/s10661-015-4990-8

Riggio, J., Baillie, J. E. M., Brumby, S., Ellis, E., Kennedy, C. M., Oakleaf, J. R., Tait, A., Tepe, T., Theobald, D. M., Venter, O., Watson, J. E. M., & Jacobson, A. P. (2020). Global human influence maps reveal clear opportunities in conserving Earth's remaining intact terrestrial ecosystems. *Global Change Biology*, *26*(8), 4344–4356. https://doi.org/10.1111/gcb.15109

Roberts, P., Hamilton, R., & Piperno, D. R. (2021). Tropical forests as key sites of the "Anthropocene": Past and present perspectives. *Proceedings of the National Academy of Sciences*, *118*(40), e2109243118. https://doi.org/10.1073/pnas.2109243118

Rodgers, W. A., & Panwar, H. S. (1988). *Panning a Wildlife Protected Area Network in India. Vol. 1.* Wildlife Institute of India.

Shahabuddin, G., & Prasad, S. (2004-04/2004-06). Assessing Ecological Sustainability of Non-Timber Forest Produce Extraction: The Indian Scenario. *Conservation and Society*, *2*(2), 235.

Singh, H. (2008). *Importance of local names of some useful plants in ethnobotanical study*. *7*(2).

Siyum, Z. G. (2020). Tropical dry forest dynamics in the context of climate change: Syntheses of drivers, gaps, and management perspectives. *Ecological Processes*, *9*(1), 25. https://doi.org/10.1186/s13717-020-00229-6

Smith, O. M., Davis, K. L., Pizza, R. B., Waterman, R., Dobson, K. C., Foster, B., Jarvey, J. C., Jones, L. N., Leuenberger, W., Nourn, N., Conway, E. E., Fiser, C. M., Hansen, Z. A., Hristova, A., Mack, C., Saunders, A. N., Utley, O. J., Young, M. L., & Davis, C. L. (2023). Peer review perpetuates barriers for historically excluded groups. *Nature Ecology & Evolution*, *7*(4), 512–523. https://doi.org/10.1038/s41559-023-01999-w

Srinivasan, U., Velho, N., Lee, J. S. H., Chiarelli, D. D., Davis, K. F., & Wilcove, D. S. (2021). Oil palm cultivation can be expanded while sparing biodiversity in India. *Nature Food*, *2*(6), 442–447. https://doi.org/10.1038/s43016-021-00305-w

Srivathsa, A., Vasudev, D., Nair, T., Chakrabarti, S., Chanchani, P., DeFries, R., Deomurari, A., Dutta, S., Ghose, D., Goswami, V. R., Nayak, R., Neelakantan, A., Thatte, P., Vaidyanathan, S., Verma, M., Krishnaswamy, J., Sankaran, M., & Ramakrishnan, U. (2023). Prioritizing India's landscapes for biodiversity, ecosystem services and human well-being. *Nature Sustainability*, *6*(5), 568–577. https://doi.org/10.1038/s41893-023-01063-2

Sterling, E. J., Filardi, C., Toomey, A., Sigouin, A., Betley, E., Gazit, N., Newell, J., Albert, S., Alvira, D., Bergamini, N., Blair, M., Boseto, D., Burrows, K., Bynum, N., Caillon, S., Caselle, J. E., Claudet, J., Cullman, G., Dacks, R., … Jupiter, S. D. (2017). Biocultural approaches to well-being and sustainability indicators across scales. *Nature Ecology & Evolution*, *1*(12), 1798–1806. https://doi.org/10.1038/s41559-017-0349-6

Stocks, G., Seales, L., Paniagua, F., Maehr, E., & Bruna, E. M. (2008). The Geographical and Institutional Distribution of Ecological Research in the Tropics. *Biotropica*, *40*(4), 397–404. https://doi.org/10.1111/j.1744-7429.2007.00393.x

Sullivan, M. J. P., Talbot, J., Lewis, S. L., Phillips, O. L., Qie, L., Begne, S. K., Chave, J., Cuni-Sanchez, A., Hubau, W., Lopez-Gonzalez, G., Miles, L., Monteagudo-Mendoza, A., Sonké, B., Sunderland, T., Ter Steege, H., White, L. J. T., Affum-Baffoe, K., Aiba, S., De Almeida, E. C., … Zemagho, L. (2017). Diversity and carbon storage across the tropical forest biome. *Scientific Reports*, *7*(1), 39102. https://doi.org/10.1038/srep39102

Tewari, V. P. (2016). Forest inventory, assessment, and monitoring, and long-term forest observational studies, with special reference to India. *Forest Science and Technology*, *12*(1), 24–32. https://doi.org/10.1080/21580103.2015.1018962

Tinkham, W. T., Mahoney, P. R., Hudak, A. T., Domke, G. M., Falkowski, M. J., Woodall, C. W., & Smith, A. M. S. (2018). Applications of the United States Forest Inventory and Analysis dataset: A review and future directions. *Canadian Journal of Forest Research*, *48*(11), 1251–1268. https://doi.org/10.1139/cjfr-2018-0196

Tiruvaimozhi, Y. V., Borah, J., Kala, C. P., Kunte, K., Patel, B., Sreejith, K. A., Thadani, R., Osuri, A. M., & Ghosh-Harihar, M. (2024). *LONG-TERM ECOLOGICAL MONITORING IN INDIA: A SYNTHESIS*. https://doi.org/10.1101/2024.08.02.606338

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., & Watson, J. E. M. (2016). Global terrestrial Human Footprint maps for 1993 and 2009. *Scientific Data*, *3*(1), 160067. https://doi.org/10.1038/sdata.2016.67

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *Dplyr: A grammar of data manipulation* [Manual].

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18