RecruitNet: A global database of plant recruitment networks

METADATA

CLASS I. Data set DESCRIPTORS

A. Data set identity:

Title: RecruitNet: A global database of plant recruitment networks

B. Data set identification code:

Data set: RecruitNet.csv and CanopyCover.csv

C. Data set description

I. Originators:

*Correspondence and requests for materials should be addressed to Miguel Verdú (Miguel.Verdu@ext.uv.es)

Miguel Verdú[1], Jose L. Garrido[2,3], Julio M. Alcántara[4,5], Alicia Montesinos-Navarro[1], Salomón Aguilar[6], Marcelo A. Aizen[7], Ali A. Al-Namazi[8], Mohamed Alifriqui[9], David Allen[10], Kristina J. Anderson-Teixeira[6,11], Cristina Armas[12], Jesús M. Bastida[13], Tono Bellido[14], Giuliano Bonanomi[15], Gustavo B. Paterno[16], Herbert Briceño[12], Ricardo A.C. de Oliveira[17], Josefina G. Campoy[18], Ghassen Chaieb[19], Chengjin Chu[20], Sarah E. Collins[1], Richard Condit[21], Elena Constantinou[22], Cihan Ü. Degirmenci[23], Leo Delalandre[24], Milen Duarte[25], Michel Faife[26], Faith Fazlioglu[27,28], Edwino S. Fernando[29,30], Joel Flores[31], Hilda Flores-Olvera[32], Ecaterina Fodor[33], Gislene Ganade[34], María Begoña Garcia[35], Patricio García-Fayos[1], Sabrina S. Gavini[7], Marta Goberna[36], Lorena Gómez-Aparicio[37], Enrique González-Pendás[38], Ana González-Robles[4], Stephen P. Hubbell[6,39], Kahraman İpekdal[40], María J. Jorquera[12], Zaal Kikvidze[41], Pınar Kütküt[23], Alicia Ledo[42], Sandra Lendínez[2], Buhang Li[43], Hanlun Liu[43], Francisco Lloret[44], Ramiro P. López[45], Álvaro López-García[2], Christopher J. Lortie[46], Gianalberto Losapio[47], James A. Lutz[48], Arantzazu L. Luzuriaga[49], František Máliš[50], Esteban Manrique[51], Antonio J. Manzaneda[4], Vinicius Marcilio-Silva[52], Richard Michalet[19], Rafael Molina-Venegas[53], José Antonio Navarro-Cano[36], Vojtech Novotny [54], Jens M. Olesen[55], Juan P. Ortiz-Brunel[56], María Pajares-Murgó[4], Nikolas Parissis[57], Geoffrey Parker[58], Antonio J. Perea[4], Vidal Pérez-Hernández[38], María Ángeles Pérez-Navarro[44], Nuria Pistón[12,59], Elisa Pizarro-Carbonell[60], Iván Prieto[12,61], Jorge Prieto-Rubio[2], Francisco I. Pugnaire[12], Nelson Ramírez[62], Rubén Retuerto[18], Pedro J. Rey[4,5], Daniel A. Rodriguez Ginart[1], Mariana Rodríguez-Sánchez[63], Ricardo Sánchez-Martín[1], Christian Schöb[49,64], Çağatay Tavşanoğlu[23], Giorgi Tedoradze[65], Amanda Tercero-Araque[4], Katja Tielbörger[66], Blaise Touzard[20], İrem Tüfekcioğlu[24], Sevda Turkis[67], Francisco M. Usero[12], Nurbahar Usta[23], Alfonso Valiente-Banuet[68,69], Alexia Vargas-Colin[31], Ioannis Vogiatzakis[22], Regino Zamora[5,70]

Affiliations

[1]Centro de Investigaciones Sobre Desertificación (CIDE, CSIC-UV-GV), Carretera de Moncada-Náquera Km 4.5, 46113, Moncada, Valencia, Spain

[2]Departamento de Microbiología del Suelo y Sistemas Simbióticos. Estación Experimental del Zaidín (EEZ-CSIC). C/ Profesor Albareda 1. 18008. Granada, Spain

[3]Estación Biológica de Doñana (EBD-CSIC). C/ Americo Vespucio 26. 41092. Sevilla, Spain

[4]Departamento de Biología Animal, Biología Vegetal y Ecología. Universidad de Jaén. E-23071 Jaén, Spain

[5]Andalusian Institute for Earth System Research (IISTA), Spain

[6]Smithsonian Tropical Research Institute (STRI) in Panama

[7]Instituto de Investigaciones en Biodiversidad y Medioambiente (INIBIOMA), Universidad Nacional del Comahue–CONICET, 8400 San Carlos de Bariloche, Argentina

[8]Life Sciences & Environment Research Institute, King Abdulaziz City for Science and Technology (KACST), 11442, Riyadh, Saudi Arabia

[9]Laboratory of Ecology and Environment, Biology Department, Faculty of Sciences Semlalia, Cadi Ayyad University, BP 2390, 40001 Marrakech, Morocco

[10]Department of Biology, Middlebury College, Middlebury VT 05753 USA

[11]Center for Conservation Ecology, Smithsonian National Zoo and Conservation Biology Institute; Front Royal, VA, USA

[12]Estación Experimental de Zonas Áridas, Consejo Superior de Investigaciones Científicas (EEZA-CSIC), Carretera de Sacramento s/n, E-04120 La Cañada, Almería, Spain

[13]Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas (EEZ-CSIC), Calle Prof. Albareda, 1, 18008, Granada, Spain

[14]Servici Devesa-Albufera, Vivers Municipals de El Saler, CV-500, km 8.5, 46012, Valencia, Spain

[15]Department of Agricultural Sciences, University of Naples Federico II, via Università 100, 80055 Portici (NA), Italy

[16]Biodiversity, Macroecology & Biogeography, University of Göttingen, Göttingen, Germany

[17]Universidade Federal do Paraná, Setor de Ciências Biológicas, Departamento de Botânica. Curitiba, PR, Brazil

[18]Departamento de Biología Funcional (Ecología), Universidade de Santiago de Compostela. E-15782 Santiago de Compostela. Spain

[19]Univ. of Bordeaux, UMR CNRS 5805 EPOC, Allée Geoffroy Saint-Hilaire – CS 50023, FR-33615 Pessac, France

[20]State Key Laboratory of Biocontrol, School of Ecology, Sun Yat-sen University, Guangzhou, 510275, China

[21]University of California, Santa Cruz, USA

[22]Faculty of Pure & Applied Sciences, Open University of Cyprus, 2252 Nicosia, Cyprus

[23]Division of Ecology, Department of Biology, Hacettepe University, Beytepe 06800, Ankara, Turkey

[24]CEFE, Univ Montpellier, CNRS, EPHE, IRD, France

[25]Instituto de Ecología y Biodiversidad (IEB), Las Palmeras 3425, Ñuñoa, Santiago, Chile

[26]Jardín Botánico de Villa Clara, Facultad de Ciencias Agropecuarias, Universidad Central 'Marta Abreu' de Las Villas, Santa Clara, Cuba

[27]Faculty of Arts and Sciences, Department of Molecular Biology and Genetics, Ordu University, 52200, Ordu, Turkey

[28]Bayreuth University (Plant Ecology, University of Bayreuth, Universitätsstraße 30, 95447, Bayreuth, Germany

[29]Institute of Biology, University of the Philippines, Diliman, Philippines

[30]Department of Forest Biological Sciences, University of the Philippines, Los Baños, Philippines

[31]Instituto Potosino de Investigación Científica y Tecnológica, A.C., División de Ciencias Ambientales, Camino a la Presa San José 2055, Col. Lomas 4a. Sección, 78216, San Luis Potosí, S.L.P., México

[32]Departamento de Botánica. Instituto de Biología, Universidad Nacional Autónoma de México, Mexico

[33]University of Oradea, Faculty of Environmental Protection, Department of Forestry and Forest Engineering, Romania

[34]Departamento de Ecología, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil

[35]Pyrenean Institute of Ecology (CSIC). Avda. Montañana 1005; 50059 Zaragoza, Spain

[36]Department of Environment and Agronomy, Centro Nacional Instituto de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Carretera de La Coruña km 7.5, 28040, Madrid, Spain

[37]Instituto de Recursos Naturales y Agrobiología de Sevilla, Consejo Superior de Investigaciones Científicas (IRNAS-CSIC), Avda. Reina Mercedes 10, E-41012 Sevilla, Spain

[38]Departamento de Investigaciones Botánicas, Centro de Investigaciones y Servicios Ambientales, Ecovida, Cuba

[39]Dept of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA

[40]Faculty of Agriculture, Ahi Evran University, Bagbasi, 40100, Kirsehir, Turkey

[41]Institute of Botany, Ilia State University, Room 310, Building F, 5 Cholokashvili Ave.Tbilisi 0162, Georgia

[42]Freelance scientist, C/Norte 20, Baros, 22712, Huesca, Spain

[43]Department of Ecology, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, 510275, China

[44]CREAF, U. Ecologia, Dept. Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma Barcelona, 08193 Cerdanyola del Valles, Spain

[45]Carrera de Biología, Facultad de Ciencias Puras y Naturales, Universidad Mayor de San Andrés (UMSA), campus universitario, calle 27, s/n, Cotacota, La Paz, Bolivia

[46]Department of Biology, York University, Toronto, ON, Canada, M3J1P3

[47]Institute of Earth Surface Dynamics, University of Lausanne, Switzerland

[48]Utah State University, Wildland Resources, 5230 Old Main Hill, Logan, Utah, 84322-5230, USA

[49]Department of Biology and Geology, Rey Juan Carlos University, E-28933. Móstoles, Madrid, Spain

[50]Technical University in Zvolen, Faculty of Forestry, T. G. Masaryka 24, Zvolen, Slovakia

[51]Real Jardín Botánico, CSIC. Plaza de Murillo 2. Madrid E-28014, Spain

[52]Department of Ecology, Evolution and Behavior, University of Minnesota, 1479 Gortner Ave, Saint Paul, MN 55108, USA

[53]Universidad de Alcalá, GLOCEE - Global Change Ecology and Evolution Group, Department of Life Sciences, Alcalá de Henares, Spain

[54]Department of Ecology and Conservation Biology, Czech Academy of Sciences, Czech Republic

[55]Department of Biology, Aarhus University, Aarhus, Denmark

[56]Departamento de Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Camino Ing. Ramón Padilla Sánchez 2100, Nextipac, Zapopan, Jalisco, 45200, Mexico

[57]Democritus University of Thrace, Department of Agricultural Development, Management of Plant Production, Plant Protection and Environment, Orestiada, Greece

[58]Smithsonian Environmental Research Center, 647 Contees Wharf Road, Edgewater MD 21037, USA

[59]Programa de Pós-graduação em Ecologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

[60]Asociación Aprisco, Torrejón el Rubio, Spain

[61]Dept. of Biodiversity and Environmental Management, Ecology Area. Fac. of Biological and Environmental Sciences, University of León. León, Spain

[62]Universidad Central de Venezuela, Facultad de Ciencias, Instituto Biología Experimental, Centro Botánica Tropical, Apartado 1041A. Caracas Venezuela

[63]Posgrado en Ciencias Biológicas, Instituto de Biología, Universidad Nacional Autónoma de México. 04510, México

[64]Institute of Agricultural Sciences, ETH Zurich, Switzerland

[65]Department of Plant Systematics and Geography, Institute of Botany, Ilia State University, Botanikuri Str. 1, 0105 Tbilisi, Georgia

[66]University of Tübingen, Institute of Evolution and Ecology, Plant Ecology Group, Auf der Morgenstelle 5, 72076 Tübingen, Germany

[67]Ordu University, Faculty of Education, Department of Mathematics and Science Education, Ordu, Turkey

[68]Departamento de Ecología de la Biodiversidad, Instituto de Ecología, Universidad Nacional Autónoma de México, AP 70-275, C.P. 04510, Ciudad de México, México

[69]Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510, México D.F., México

[70]Department of Ecology, University of Granada, Spain

II. Abstract:

Plant recruitment interactions (i.e., who recruits under whom) shape the composition, diversity and structure of plant communities. Despite the huge body of knowledge on the mechanisms underlying recruitment interactions between species, we still know little about the structure of the recruitment networks emerging in ecological communities. Modeling and analyzing the community-level structure of plant recruitment interactions as a complex network can provide relevant information on ecological and evolutionary processes acting both at the species and ecosystem levels. We report a data set containing 143 plant recruitment networks in 23 countries across five continents, including temperate and tropical ecosystems. Each network identifies the species under which another species recruits. All networks report the number of recruits (i.e., individuals) per species. The data set includes >850,000 recruiting individuals involved in 118,411 paired interactions between 3,318 vascular plant species across the globe. The cover of canopy species and open ground is also provided. Three sampling protocols were used: 1) The Recruitment Network (RN) protocol (106 networks) focuses on interactions between established plants ("canopy species") and plants in their early stages of recruitment ("recruit species"). A series of plots are delimited within a locality and all the individuals recruiting and their canopy species are identified; 2) The paired Canopy-Open (pCO) protocol (26 networks) consists in locating a potential canopy plant and identifying recruiting individuals under the canopy and in

a nearby open space of the same area; 3) The Georeferenced plot (GP) protocol (11 networks) consists in using information from georeferenced individual plants in large plots to infer canopy-recruit interactions. Some networks incorporate data for both herbs and woody species, while others focus exclusively on woody species. The location of each study site, geographical coordinates, country, locality, responsible author, sampling dates, sampling method and life habit of both canopy and recruit species are provided. This database will allow researchers to test ecological, biogeographical and evolutionary hypotheses related to plant recruitment interactions. There are no copyright restrictions on the data set; please cite this data paper when using these data in publications.

D. Key words: Ecological networks; Facilitation; Plant-plant interactions; Recruitment; Replacement.

CLASS II. RESEARCH ORIGIN DESCRIPTORS

A. Overall project description

1. **Identity:** Data on plant recruitment interactions shaping ecological networks in communities all over the world.

2. **Originators:** M. Verdú, JL. Garrido, JM. Alcántara and A. Montesinos-Navarro conceived and coordinated the project. All coauthors provided data on plant recruitment networks that were finally assembled in the data set by JL Garrido.

3. **Period of Study:** Data collection spans from 2006 to 2021.

4. **Objectives:** Provide a worldwide georeferenced database of plant recruitment interactions at the community level that can be used for testing ecological and evolutionary hypotheses related to community dynamics, interaction specificity, biogeographical trends, etc. across broad spatial scales.

5. Abstract

Plant recruitment interactions (i.e., who recruits under whom) shape the composition, diversity and structure of plant communities. Despite the huge body of knowledge on the mechanisms underlying recruitment interactions between species, we still know little about the structure of the recruitment networks emerging in ecological communities. Modeling and analyzing the community-level structure of plant recruitment interactions as a complex network can provide relevant information on ecological and evolutionary processes acting both at the species and ecosystem levels. We report a data set containing 143 plant recruitment networks in 23 countries across five continents, including temperate and tropical ecosystems. Each network identifies the species under which another species recruits. All networks report the number of recruits (i.e., individuals) per species. The data set includes >850,000 recruiting individuals involved in 118,411 paired interactions between 3,318 vascular plant species across the globe. The cover of canopy species and open ground is also provided. Four sampling protocols were used: 1) The Recruitment Network (RN) protocol (106 networks) focuses on interactions between established plants ("canopy species") and plants in their early stages of recruitment ("recruit species"). A series of plots are delimited within a locality and all the individuals recruiting and their canopy species are identified; 2) The paired Canopy-Open (pCO) protocol (26 networks) consists in locating a potential canopy plant and identifying recruiting individuals under it and in a nearby open space of the same area; and 3) The Georeferenced plot (GP) protocol (11 networks) consists in using information from georeferenced individual plants in large plots to infer canopy-recruit interactions. Some networks incorporate data for both herbs and woody species, while others focus exclusively on woody species. The location of each study site, geographical coordinates, country, locality, responsible author, sampling dates, sampling method and sample size are provided. This database will allow researchers to test ecological, biogeographical and evolutionary hypotheses related to plant recruitment interactions. There are no copyright restrictions on the data set; please cite this data paper when using these data in publications.

6. **Sources of funding:** The coordination of this data set was supported by projects PGC2018-100966-B-100, PID2020-113157GB-I00 (funded by MCIN/AEI/10.13039/501100011033 and "ERDF A way of making Europe"), CYTED (Red 418RT0555), GVA (CIPROM/2021/63) and SUMHAL (LIFEWATCH-2019-09-CSIC-13, POPE 2014-2020). Storage and handling of recruitment networks datasets was hosted in the eLab "Recruitment/replacement networks" provided by the Singular Scientific-Technical Infrastructure (ICTS-RBD) of the Estación Biológica de Doñana (EBD-CSIC) (MICINN FEDER Funds ICTS-2017-08-CSIC-4).

B. Specific subproject description

1. Site description

a. Site type: A wide range of terrestrial plant communities, including forests, shrublands, and grasslands.

b. Geography: Worldwide (Figure 1)

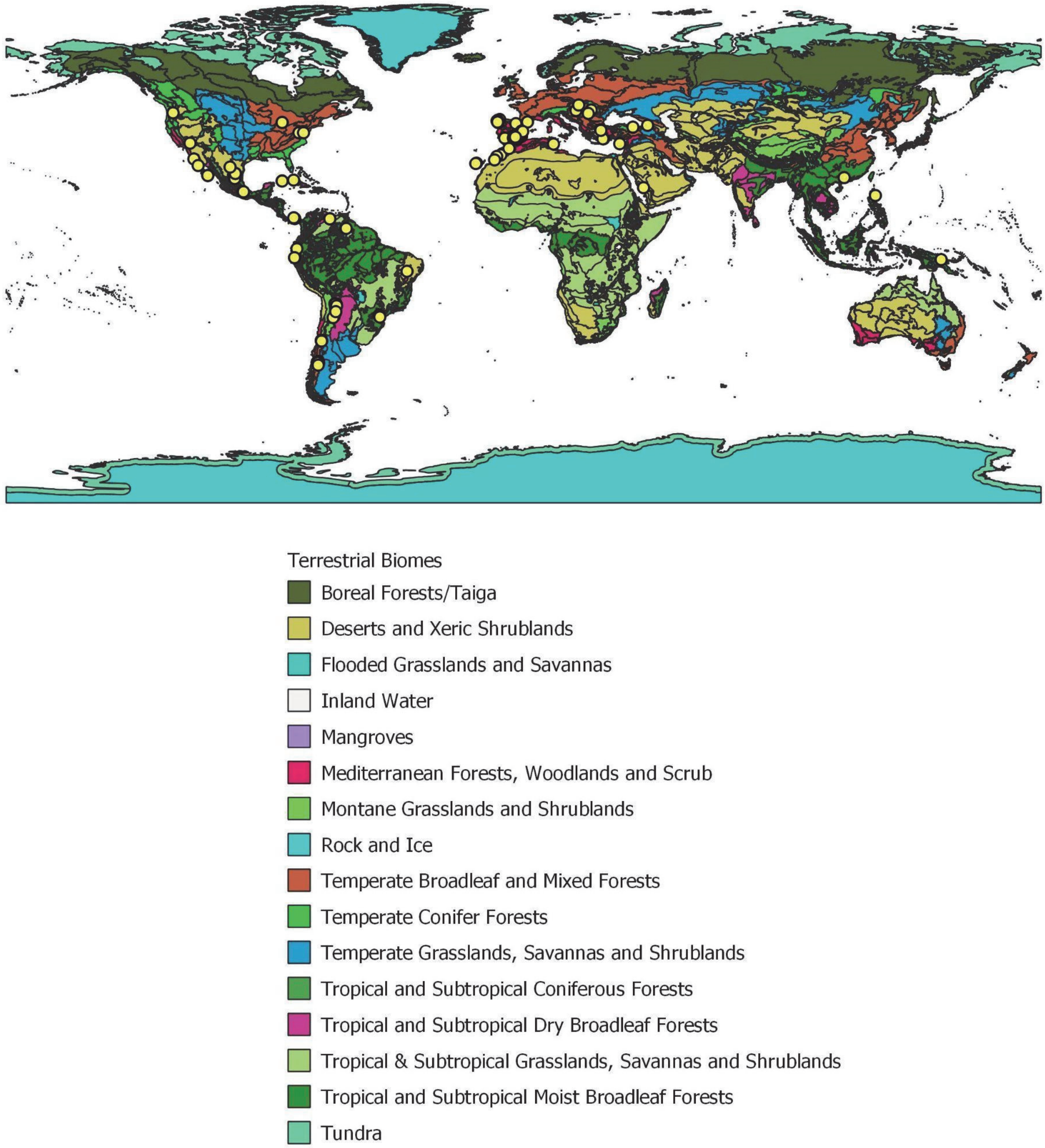

Figure 1. Map of the study sites where recruitment networks were sampled across different biomes, as categorized by Olson et al. (2001).

c. Habitat: A wide range of terrestrial habitats that are named in each record of the data set including additional information about disturbance history.

2. Sampling design and Research methods

We compiled data describing plant-plant recruitment networks, i.e., which species recruit under which others in the whole plant community. A recruit individual is defined as a young, non-reproductive individual that is typically less than one fourth the size of a fully grown adult of the species (Alcántara et al. 2019a). A canopy individual is the plant more directly conditioning the microenvironment of the recruit both above and belowground. In practical terms, a canopy individual is one that houses the recruit under its canopy. When a recruit is located under multiple canopy layers, as usually occurs in tropical forests, we consider the canopy plant to be the nearest adult plant whose trunk is located less than a distance *d*. The precise values of *d* must be adapted to the particular habitat and canopy species, so they are based on expert knowledge. In any case, such distance should ensure that the roots of the canopy individuals are in contact with those of the recruit individuals (see Alcántara et al. 2019a for further details). As an example, *d* was considered to be 0.5 m in Mediterranean pine–oak forests (Alcántara et al. 2019b).

Studies focusing on a small subgroup of the canopy species present in the community were discarded because they did not capture the whole network of interactions occurring in the community. We also excluded plant-plant interaction networks based on co-occurrence of adults (e.g., Saiz and Alados 2011, Saiz et al 2018; Soliveres and Maestre 2014) because of the impossibility to discern between the canopy and the recruit species. All networks included in the data set followed one of the three protocols described in Figure 2.

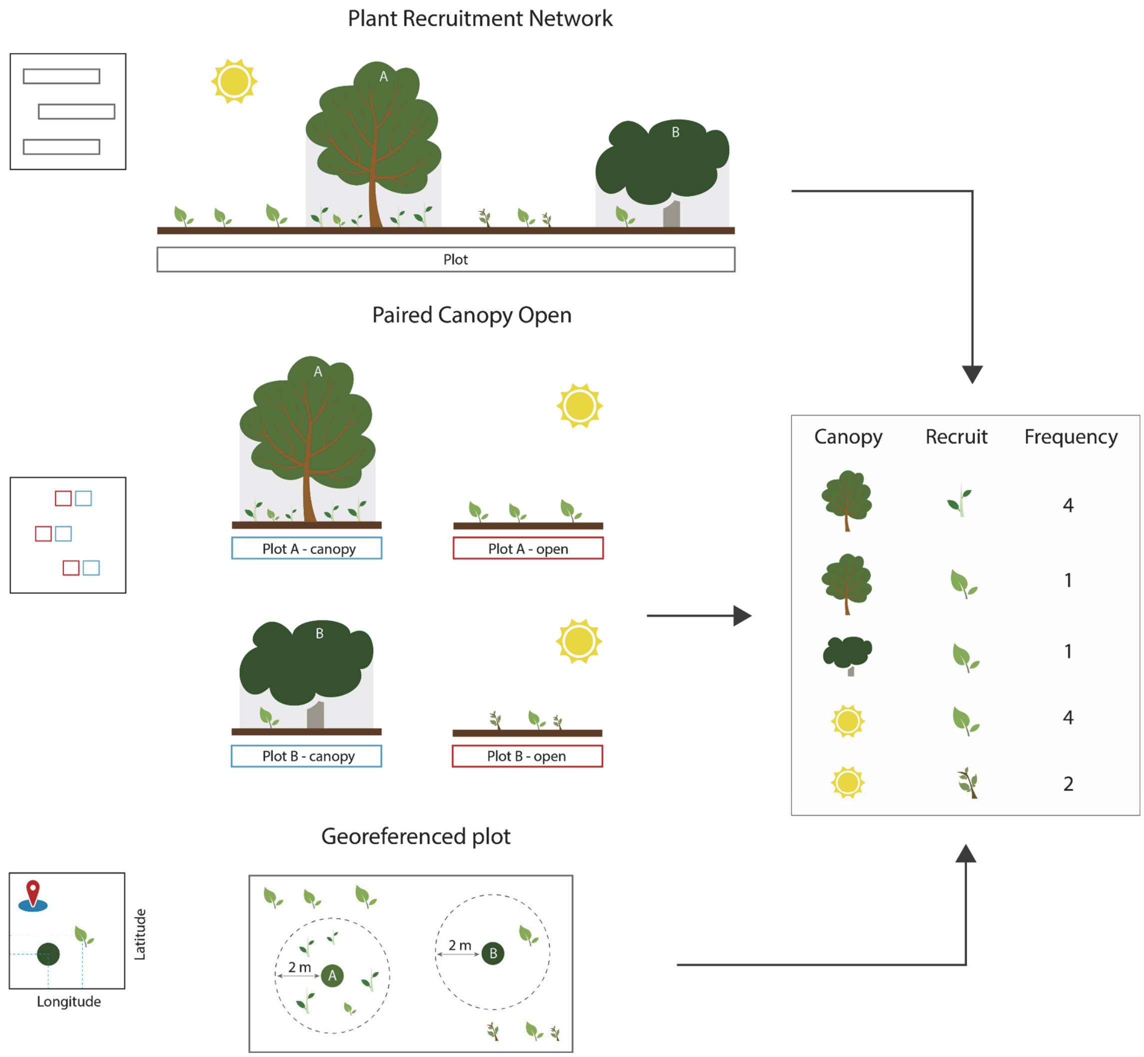

Figure 2. Three sampling protocols (left drawings) lead to the same recruitment network (interaction matrix in the right). The **Recruitment Network (RN) protocol** (106 networks in the database) focuses on interactions between established plants (called "canopy species") and plants in their early life stages (called "recruit species"). A series of sampling units of fixed area (e.g., quadrats, plots or transects) are delimited within a locality and all the individuals recruiting within the area and their canopy species are identified. Each sampling unit is large enough to be considered as a replicate of the network corresponding to the studied community (e.g., it could potentially contain at least one adult individual of each of the species in the community). The **paired Canopy-Open (pCO) protocol** (26 networks) consists in locating a potential canopy individual and identifying individual plants recruiting beneath it and in a nearby open space of the same area (see for example Rey et al. 2016). Here, each sampled pair is a replicate of the possible links involving a given canopy species or the open node, so all potential canopy plants in the plot must be sampled in order to obtain the full network of the studied community. This protocol is typically used in studies specifically designed to determine the importance of facilitation for recruitment, when canopy species are

previously identified (Navarro-Cano et al., 2019). The **Georeferenced plot (GP) protocol** (11 networks) takes advantage of available information from mapped individual plants in large plots. To transform this type of data into canopy-recruit interactions we made a series of assumptions regarding the size of recruits of each species and the maximum distance between canopy and recruited plant. Basically, we considered as a recruit any individual with DBH lower than 10% that of the largest conspecific present in the plot. Besides, we considered that a recruit was interacting with the nearest adult plant located less than 2 m away, and considered that it was recruiting in an open interspace if there were no adult plants less than 2 m from the recruit. In all protocols, the network corresponding to the studied community should be built by aggregating the information from all the sampled plots or pairs. Nevertheless, we provide the information disaggregated because it reflects the actual sampling scheme and because it can be useful, for example, in studies exploring network properties at different spatial scales or the spatial variability of the interactions within communities. All icons in this Figure are adapted from the Noun Project under a Creative Commons license CC-BY 3.0; Tree A by Aleksandr Vector (https://thenounproject.com/icon/trees-1009161/), Tree B by Lance Hancock (https://thenounproject.com/icon/bush-40537/), Sapling 1 by Vectors Market (https://thenounproject.com/icon/sapling-1931834/), Sapling 2 by Mr Balind ID (https://thenounproject.com/icon/seeds-1168067/), Sapling 3 by Domingo Marti (https://thenounproject.com/icon/sapling-653060/), Sun by Muammar Khalid (https://thenounproject.com/icon/sun-1342344/), GPS icon by Dinosoftlab (https://thenounproject.com/icon/gps-1992295/).

Despite the different assumptions made by each protocol, all of them should lead to the same network of interactions, as depicted in Figure 2. To check whether the three protocols have differently sampled the plant recruitment interactions, we ran a coverage analysis. We obtained abundance-based estimates of network coverage from the observed frequency of each interaction present in a recruitment network. Estimates of interaction coverage followed the procedure proposed by Chao et al. (2014) to assess species richness and were obtained using the R package iNEXT (Hsieh et al. 2016). Link coverage averaged 0.95, ranging between 0.60 and 1. Most studies (86.71%) reached high link coverage (>0.90), although there were cases of relatively low coverage in studies using any sampling protocol (Fig. 3). There was a trend of decreasing coverage with increasing network size (Table 1), although the decrease in coverage with network size should be a concern only for the largest networks containing more than 200 nodes (Fig. 3).

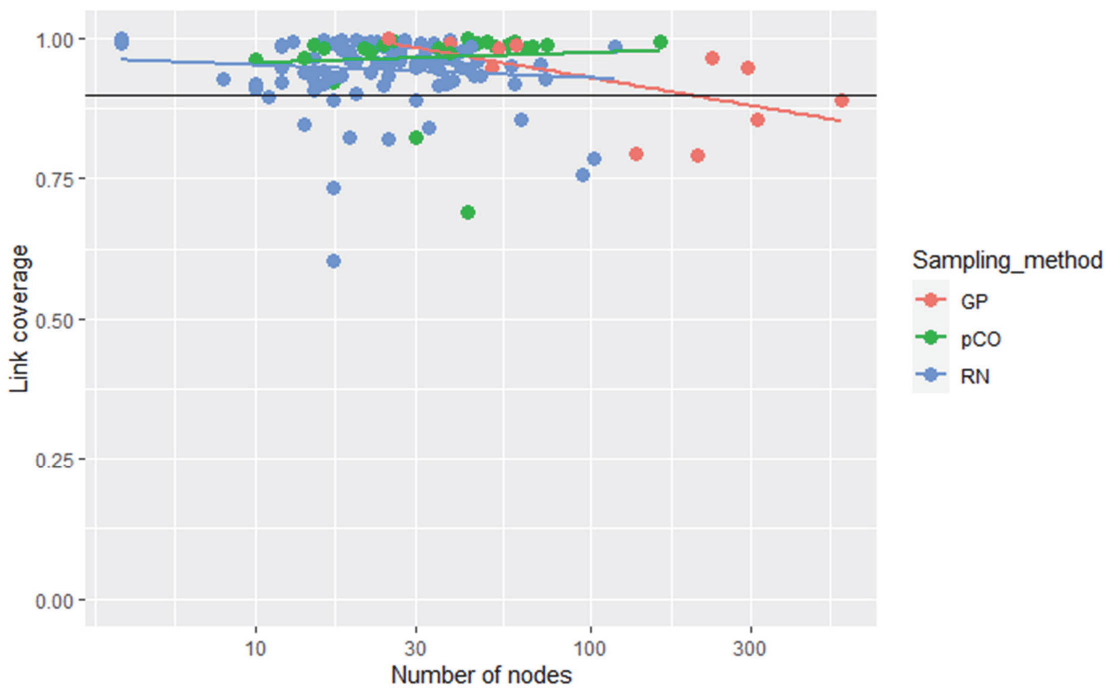

Figure 3. Relationship between link coverage and network size (number of nodes). Different colors indicate different sampling protocols. The black horizontal line marks a coverage of 0.9. Note the base 10 logarithmic scale in the x-axis.

Table 1. Results of Generalized Linear Model testing the effect of sampling method, network size (log-transformed number of nodes) and proportion of herb species included in the network on estimated link coverage. The model was fitted with beta family distribution using the R package glmmTMB (Magnusson et al. 2017).

Effect	Chi Square	df	P-value
Sampling method	3.11	2	0.21
Number of nodes	9.71	1	0.002
Proportion of herbs	1.13	1	0.29
Sampling method x Number of nodes	1.61	2	0.45
Sampling method x Proportion of herbs	3.87	2	0.14

The database contains information for 143 recruitment networks from all vegetated biogeographic realms (Table 2A), although Australasia, the Indomalayan and Afrotropical realms are very scarcely represented. This biogeographical limitation is common in studies of all types of ecological networks (Poisot et al. 2021). RecruitNet includes networks from 48 ecoregions of the world (Olson et al. 2001), representing 10 of the 14 types of terrestrial biomes (Table 2B). Regarding the successional stage, 5% correspond to early successional communities, 13% to mid successional ones, 78% to late successional and 4% lacked information on this regard.

We also provide data on the growth form of canopy and recruit species through a simple classification of species as woody or herb. This information is actually a methodological difference between the networks (some networks do not contain herbs because the original research focused only on woody plants, not because herbs are absent in the studied vegetation).

The number and size of plots are provided for each network in the data set. The studies vary in their sampling effort depending on the type of community studied, using between 1 and 1156 plots, and plot areas between 0.04 (for plots focused on herbs) and 500,000 m^2 (in tropical forest plots). The area sampled in a given community ranged between 1.65 and 500,000 m^2, with a median of 1758 m^2. The networks had a median of 26 nodes and 106 links, ranging between a minimum of 4 nodes and 7 links in fir forests from Turkey, and a maximum of 527 nodes and 40,365 links in the forest plot of Wanang (Papua-New Guinea).

Table 2. Summary of the distribution of networks available in RecruitNet across (A) biogeographic realms, countries and (B) biomes (classified according to Olson et al. 2021).

A) Biogeographic Realm	Countries (Number of networks)
Palearctic	Spain (56), Turkey (5), Morocco (4), Romania (2), Slovakia (2), Cyprus (1), Georgia (1), Tunisia (1)
Indomalayan	China (1), Philippines (1)
Nearctic	Mexico (14), USA (5)
Neotropical	Argentina (12), Bolivia (8), Chile (7), Mexico (6), Venezuela (4), Brazil (3), Panama (3), Cuba (3), Ecuador (1), Peru (1)
Afrotropical	Saudi Arabia (1)
Australasian	Papua New Guinea (1)
B) Number of networks	**Biomes represented in RecruitNet**

22	Deserts and Xeric Shrublands
58	Mediterranean Forests, Woodlands and Scrub
29	Temperate Broadleaf and Mixed Forests
3	Temperate Conifer Forests
7	Montane Grasslands and Shrublands
6	Tropical and Subtropical Grasslands, Savannas and Shrublands
8	Tropical and Subtropical Dry Broadleaf Forests
1	Tropical and Subtropical Coniferous Forests
7	Tropical and Subtropical Moist Broadleaf Forests
2	Mangroves

To facilitate future studies, we provide standardized species names following World Flora Online (WCVP, 2022) and Missouri Botanical Gardens Tropicos databases (Tropicos.org, 2022) as references. Taxonomic standardization with partial matching was performed with the function TNRS() from the R package TNRS by Maitner and Boyle (2022). Taxonomic affiliation to the family level is also provided to help future analyses. After the standardization, we report records from 1362 different genus and 215 families. The great majority of names were at the species level but some names were also at higher (subspecies, variety) or lower taxonomic resolution (genus, family). In the case of unidentified species, the standardized identification was retained as the original one (e.g., unidentified_sp1). The most represented families were Fabaceae, Asteraceae, Rubiaceae, Lamiaceae and Euphorbiaceae (Fig. 4).

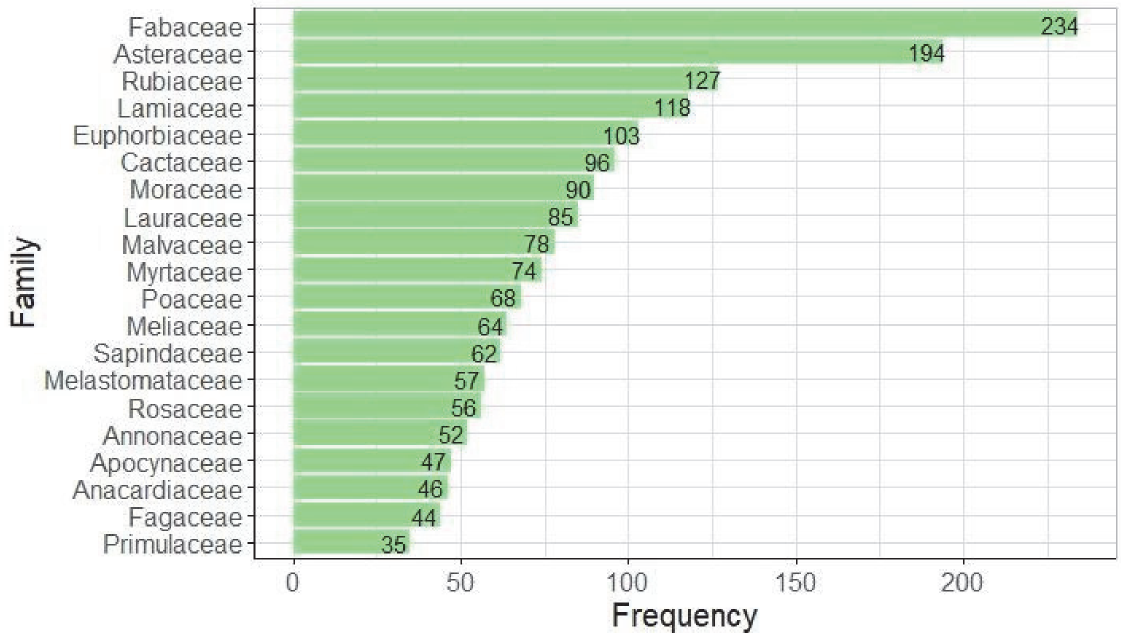

Figure 4. Number of species within the most represented families in the database.

Disturbance types identified in each site (Disturbance_ID) were categorized in (1) human disturbances, i.e., those derived from direct human activities or domestic animals; (2) climatic or stochastic events, i.e., eventual natural disturbances that may severely affect plant communities; (3) none, i.e., when no disturbances were observed or were expected to occur in natural systems (e.g., herbivory); and NA, when disturbances were not recorded. The original disturbance identification was also retained.

The cover of each canopy species and the open ground were provided as a separate dataset to facilitate the incorporation of null models controlling for the microhabitat cover in future studies. Cover data for RN and pCO networks were estimated through the line-intercept method (Canfield, 1941) or directly recording the canopy projection of all the canopy individuals in the plot (Alcántara et al. 2019b). To estimate the cover of a canopy plant in GP (which we define as the area around the trunk such that a recruiting individual would be interacting with the canopy individual), we first draw the Voronoi polygon of each individual plant in the plot and the circle of 2 m radius around them. The cover of an individual was assumed to be the area of the intersection between its Voronoi polygon and its 2 m buffer. Voronoi polygons have been successfully used to predict tree crown areas (Goudie et al. 2009) and to simulate the natural dynamics of forest stands with closed canopies (Abellanas et al. 2016). Finally, the cover of each species in the plot is the sum of the cover of all individuals of the same species.

The data collected to construct recruitment networks can be viewed as a snapshot of the status of the community. However, a network approach can improve our understanding of plant community dynamics (Alcántara et al. 2015). Recruitment networks are matrices describing the frequency of recruitment of a given species under the canopy of every other species in the community. Therefore, these matrices contain the raw data to develop Markov chain models of community dynamics (Siles et al. 2008). This is because plant community dynamics can be modelled as a process where individuals recruiting under a plant can replace it after its death (e.g. Horn 1975). For example, by applying a Markov transitional approach to recruitment matrices, Siles et al (2008) predicted post-fire vegetation dynamics and assessed the long-term contribution of canopy plants to forest restoration. Adding a temporal dimension to the recruitment networks would provide invaluable information about successional dynamics. This database represents an outstanding opportunity to resample the networks in the future and monitor changes in the structure of plant communities.

C. Project personnel: Coauthors of this data paper.

CLASS III. data set STATUS AND ACCESSIBILITY

A. Status

1. **Latest update:** April 2022.
2. **Latest archive date:** April 2022.
3. **Metadata status:** Last update on April 2022.
4. **Data verification:** We checked all the information such as species names, geographic coordinates, and percentage cover calculation. We conserved the taxonomic nomenclature used by each contributor and provided accepted names according to World Flora Online (WCVP, 2022) and Missouri Botanical Gardens Tropicos databases (Tropicos.org, 2022).

B. Accessibility

1. Storage location and medium: The complete dataset is available as Supporting Information and associated data is also available at Zenodo repository (https://doi.org/10.5281/zenodo.6567608) and cited as Verdú et al. (2022).

2. Contact person: Miguel Verdú (Miguel.Verdu@ext.uv.es). Centro de Investigaciones sobre Desertificación (CSIC-UV-GV), Carretera Moncada-Náquera Km 4,5. 46113 Moncada, Valencia (Spain). Tel: +34 963424204. Fax: +34 963424160.

3. Copyright restrictions: None.

4. Proprietary restrictions: Please cite this data paper when using the current data in publications.

C. Costs

None.

CLASS IV. DATA STRUCTURAL DESCRIPTORS A.

Dataset File

1. **Identity:** RecruitNet.csv and CanopyCover.csv
2. **Size:** RecruitNet.csv: 135,211 rows and 26 columns, with 135,210 records stored in 39,515 MB.

   CanopyCover.csv: 26,812 rows and 7 columns, with 26,2811 records stored in 1,437 KB.
3. **Format and storage mode:** comma-separated values (.csv)
4. **Header information:** See column descriptions in section B.
5. **Alphanumeric attributes:** Mixed.

B. Variable information

The information regarding the variables contained in the 26 columns of the database describing the plant recruitment interactions (RecruitNet.csv) and in the 7 columns of the database containing the cover of the canopy plants (RecruitCover.csv) are described in Table 3.

Table 3. Names and description of the variables contained in the databases RecruitNet.csv and CanopyCover.csv.

RecruitNet.csv

Variable name	**Descriptor**
Study_site	ID for each network
Location	Location of the study site
Site_responsible	Name of the principal investigator collecting the data
Country	Name of the country where the network was sampled
Biome	Biome defined as Major Habitat Type according to Olson et al. (2021)
Vegetation_type	Type of vegetation where the network was sampled
Community	Name of the plant community
Successional_stage	Early, mid or late successional stage
Disturbance	Type of disturbance (eg., browsing, fire, etc)
Disturbance_ID	Identification for disturbance types (1) human disturbances; (2) climatic or stochastic events; (3) none
Latitude	Latitude in decimal degrees
Longitude	Longitude in decimal degrees
PlotdimX	Dimension X of the plot in meters
PlotdimY	Dimension Y of the plot in meters
Sampling_date	Sampling date (year)
Plot	ID of the plot within the study site

Canopy	Scientific name of the canopy species or “Open”, indicating open ground.
Recruit	Scientific name of the recruiting species
Standardized_Canopy	Standardized name of the canopy species according to World Flora Online and Missouri Botanical Gardens Tropicos databases
Standardized_Recruit	Standardized name of the recruit species according to World Flora Online and Missouri Botanical Gardens Tropicos databases
Family_Canopy	Taxonomic family of the canopy species
Family_Recruit	Taxonomic family of the recruit species
Frequency	Number of individuals of the recruit species growing under individuals of the canopy species within a plot.
Sampling_method	Recruitment Network (RN); paired Canopy-Open (pCO); Georeferenced Plots (GP)
LifeHabit_Canopy	Life habit of the canopy species (Woody, Herb, variable)
LifeHabit_Recruit	Life habit of the recruit species (Woody, Herb, variable)

CanopyCover.csv

Variable name	**Descriptor**
Study_site	ID for each network
Plot_ID	ID of the plot within the study site
Canopy	Scientific name of the canopy species or “Open”, indicating open ground.
Cover	Cover percentage of canopy species (or open ground)

Sampled_distance_or_area	Distance (m) or area (m^2) sampled to estimate canopy cover when the Method_cover is “line” or “plot” respectively
Sampling_method	Recruitment Network (RN); paired Canopy-Open (pCO); Georeferenced Plot (GP)
Method_cover	Method employed to sample canopy cover; “line” when the line-intercept method (Canfield, 1941) was used or "plot" when cover were directly recorded the canopy projection of all the canopy individuals in the plot (Alcántara et al. 2019b).

C. Data anomalies: Missing data were coded as “NA”.

CLASS V. SUPPLEMENTAL DESCRIPTORS

A. Data acquisition

1. Data acquisition methods: Field data were collected by experienced botanists and ecologists that coauthored this data paper. The networks from Barro Colorado Island (BCI8), Cocoli, Sherman, Michigan Big Woods, Palanan, SCBI, SERC, Wanang and Wind River, were extracted with permission from the responsible researchers from data available in Forest Geo Global Earth Observatory (see forestgeo.si.edu for further information of these plots).

2. Data entry verification procedures: Data sets were required to fit a specific spreadsheet format that was finally curated and supervised by JL Garrido.

B. Quality assurance/quality control procedures: The coordination team visually checked that all data matched the quality requirements.

C. Related materials: None.

D. Computer programs and data-processing algorithms:

Individual data sets were formatted in two csv files with the help of the R-software (https://www.r-project.org/).

E. Archiving

1. Archival procedures: All networks and cover datasets were saved in two separated csv files.

2. Redundant archival sites: None.

F. Publications and results

None.

G. History of data set usage

1. Data request history: None.

2. Data set update history: None.

References

Abellanas, B., Abellanas, M., Pommerening, A., Lodares, D., and Cuadros, S. 2016. "A forest simulation approach using weighted Voronoi diagrams. An application to Mediterranean fir *Abies pinsapo* Boiss stands". *Forest Systems* 25: e062-e062.

Alcántara, J. M., Rey, P. J., and Manzaneda, A. J. 2015. "A model of plant community dynamics based on replacement networks". *Journal of Vegetation Science* 26: 524-37.

Alcántara, J.M., Garrido, J.L., Montesinos-Navarro, A., Rey, P. J., Valiente-Banuet, A. and Verdú, M. 2019a. "Unifying facilitation and recruitment networks." *Journal of Vegetation Science* 30: 1239-49.

Alcántara, J. M., Garrido, J. L., and Rey, P. J. 2019b. "Plant species abundance and phylogeny explain the structure of recruitment networks". *New Phytologist* 223: 366-376.

Canfield, R.H. 1941. "Application of the line interception method in sampling range vegetation." *Journal of Forestry* 39: 388-394.

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., and Ellison, A. M. 2014. “Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies.” *Ecological Monographs* 84: 45-67.

Goudie, J. W., Polsson, K. R., and Ott, P. K. 2009. “An empirical model of crown shyness for lodgepole pine (*Pinus contorta* var. *latifolia* [Engl.] Critch.) in British Columbia”. *Forest Ecology and Management* 257: 321-331.

Hsieh, T. C., Ma, K. H., and Chao, A. 2016. “iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers).” *Methods in Ecology and Evolution* 7: 1451-56.

Horn, H.S. 1975. “Markovian properties of forest succession.” in M.L. Cody and J.M. Diamond (editors), Ecology and Evolution of Communities, pp. pages 196-211. Harvard University Press.

Magnusson, A., Skaug, H., Nielsen, A., Berg, C., Kristensen, K., Maechler, M., van Bentham, K. J., Bolker, B. M. and Brooks, M. M. 2017. “glmmTMB: Generalized Linear Mixed Models using Template Model Builder. R package version 0.1.3.” https://github.com/glmmTMB.

Maitner, B. and Boyle, B. 2022. “TNRS: Taxonomic Name Resolution Service. R package version 0.3.0.”

Navarro-Cano, J., Goberna, M. and Verdú, M. 2019. “Using plant functional distances to select species for restoration of mining sites.” *Journal of Applied Ecology* 56: 2353-62.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., ... and Kassem, K. R. 2001. “Terrestrial Ecoregions of the World: A New Map of Life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity.” *BioScience* 51: 933-8.

Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., Mercier, B., Violet, C. and Vissault, S. 2021. "Global knowledge gaps in species interaction networks data." *Journal of Biogeography* 48: 1552-63.

Rey, P. J., Alcántara, J. M., Manzaneda, A. J., and Sánchez-Lafuente, A. M. 2016. “Facilitation contributes to Mediterranean woody plant diversity but does not shape the diversity–productivity relationship along aridity gradients.” *New Phytologist* 211: 464-76.

Saiz, H., and Alados, C. L. 2011. “Structure and spatial self-organization of semi-arid communities through plant–plant co-occurrence networks.” *Ecological Complexity* 8: 184-91.

Saiz, H., Gómez-Gardeñes, J., Borda, J. P., and Maestre, F. T. 2018. “The structure of plant spatial association networks is linked to plant diversity in global drylands.” *Journal of Ecology* 106: 1443-53.

Siles, G., Rey, P.J., Alcántara, J.M. and Ramírez, J.M. 2008. “Assessing the long-term contribution of nurse plants to restoration of Mediterranean forests through Markovian models.” Journal of Applied Ecology 45: 1790–98.

Soliveres, S., and Maestre, F. T. 2014. “Plant–plant interactions, environmental gradients and plant diversity: A global synthesis of community-level studies.” *Perspectives in Plant Ecology, Evolution and Systematics* 16: 154–63.

Tropicos.org. 2022. Missouri Botanical Garden. 01 Mar 2022 https://tropicos.org

Verdú et al (2022). Datasets on plant recruitment networks. Zenodo. https://doi.org/10.5281/zenodo.6567608

WCVP (2022). “World Checklist of Vascular Plants, version 2.0. Facilitated by the Royal Botanic Gardens, Kew.” Published on the Internet; http://wcvp.science.kew.org/ Retrieved 01 March 2022.