Biodesert facilitation data

2 March 2024

Attribute	Explanation		Notes
File names	v4_v3_inter_algeria.csv, v4_v3_inter _argentina.csv, v4_v3_inter _australia.csv, v4_v3_inter _chile.csv, v4_v3_inter _chinachong.csv, v4_v3_inter _chinaxin.csv, v4_v3_inter _iranabedi.csv, v4_v3_inter _iranfarzam.csv, v4_v3_inter _israel.csv, v4_v3_inter _namibiablaum.csv, v4_v3_inter _namibiawang.csv, v4_v3_inter _southafrica, v4_v3_inter _spainmaestre.csv, v4_v3_inter _spainrey.csv		The data from each country is called v4_v3_inter _countryname.csv. Where more than one dataset was contributed from the same country, the surname of the contributed was added to the country's name.
File location	Biodesert facilitation data for sharing > Countriesv3		
File type	.csv file with semicolon (;) as separator		
File origin	Raw data was received from Victoria Ochoa and each country's data was exported to a separate .csv file and cleaned separately. Exception: the Chinaxin, Namibiawang and Iranfarzam data used in cleaning come from the collaborators themselves after they applied certain corrections. Data was cleaned and standardised in R with the script "data and cleaning and standardising", see below for changes made during data cleaning.		Data cleaning was completed in September 2023.
Analyst	Imke Chrissie Smit, MSc candidate at the University of Pretoria, South Africa. Contact information: u19053721@up.ac.za or imke610@gmail.com.		The analyst did the data cleaning and compiled this metadata file.
Variables	ID	Unique ID number for each plot.	
	COU	Name of country.	
	RES	Researcher in charge of collecting the data.	
	SITE	Site name.	
	SITE_ID	Unique ID number of the site.	
	PLOT	Original Lab code for grazing intensity within the plot (1:high grazing; 2:medium grazing; 3:low grazing; 4: ungrazed).	
	Date	date of the facilitation survey.	
	Microsite	1 indicates bare soil, 2 indicates a nurse plant.	
	ID_Microsite	name of the nurse plant.	
	diameter_1.cm.	diameter 1 of the nurse plant.	
	diameter_2.cm.	diameter 2 of the nurse plant.	
	canopy.area..cm2.	Area of the patch, assuming a spherical shape (only for plant microsites).	
	Number.of.replicate	Unique replicate number of a nurse-open pair in a plot.	Different nurse species in the same plot do not

			share replicate numbers. E.g., reps 1-25 assigned to nurse 1, reps 26-50 assigned to nurse 2, reps 51-75 assigned to nurse 3.
	Quadrat	Number of quadrats placed within the replicate to make up the area of the nurse.	Do not use this in analysis. Some Quadrats were lost when averaging species cover the number of Quadrats.
	Species.within.quadrat	Name of the species growing in the nurse or bare microsite.	NA means no species were found in the microsite.
	Cover	% cover of the species in the microsite.	NA means no cover was recorded for the species or no species were found in the microsite.
	Number.of.individuals	Abundance of species in the microsite.	NA means no abundance was recorded for the species or no species were found in the microsite.
	microsite_rep	A unique code that was created for each microsite by concatenating Site-ID, PLOT, ID_Microsite and Number of replicate.	
	GRAZ	Grazing intensity within the plot (0 ungrazed; 1:low grazing; 2:medium grazing; 3:high grazing).	Source:BIODESERT_sites _information.xlsx
	ARIDITY.v3	Aridity value of a plot.	Source:BIODESERT_sites _information.xlsx

Data cleaning and standardising

Extensive data cleaning and standardising was necessary before analyses could be done. Contributors were contacted about problems in their data. The following changes were applied to all facilitation datasets:

- Empty rows were removed.
- A unique code was created for each microsite by concatenating Site-ID, PLOT, ID_Microsite and Number of replicate. This code is called the microsite_rep.
- If a microsite_rep had no species (as denoted by a space, dash, "empty", NaN or "Nada" in "Species within quadrat", the entry in "Species within quadrat" was set to NA.
- If present in a dataset, any entries in "Species within quadrat" referring to moss, bryophytes or cryptobiotic soil crust were removed. If this was the only entry in a microsite_rep, the entry was changed to NA. If it was not the only entry in a microsite_rep, the row was deleted.

- The "ID_Microsite" variable was standardised to have "Bare" indicating the microsite without the dominant species.
- The Microsite variable was standardised so that a value of 1 denotes the bare microsite and a value of 2 denotes the nurse microsite.
- Nurse and species names with multiple spellings were standardised to one spelling.
- Infraspecies ranks were removed.
- All species names were changed to match those in the functional trait and quadrat dataset. The rules for this were as follows:
  - Infraspecies ranks were removed in the facilitation and the trait data so that species match.
  - Names were not matched if only the genus matched. E.g. "*Thymus*" in the trait data and *Thymus vulgare* in the trait data were not regarded as the same.
  - Names with sp. or indet. were regarded as the same. E.g *Thymus* sp. in the facilitation data were regarded the same as *Thymus* indet. in the trait data.
  - Names with cf were regarded as the same. E.g. cf. *Rhuschia spinosa* in the facilitation data were regarded the same as *Rhuschia spinosa* in the trait data.
- If applicable, replicate numbers within a plot were changed so that the replicate numbers of each nurse followed on from the previous nurse. E.g., reps 1-25 assigned to nurse 1, reps 26-50 assigned to nurse 2, reps 51-75 assigned to nurse 3.

The following changes were specific to the data submitted by each group:

Algeria
- Site 1 plot 3 replicate 7 contained only a Bare microsite. The replicate number in the second instance of 1-3-Stipa tenacissima-6 was renamed to be included in replicate 7 (thus the microsite_rep was renamed to 1-3-Stipa tenacissima-7).

Argentina
- The microsite variable was overwritten so that 1 corresponds to a bare microsite and 2 corresponds to a nurse microsite.
- Site 7 plot 2 replicate 52 had only nurse microsites. The replicate number of 7-2-Bare-53 was changed to 52. And the replicate number of 7-2-Bare-54 was changed to 53.
- Replicate numbers in site 6 plot 1 were changed so that the replicates of each nurse followed on from the previous nurse.
- A row containing "" in "Species within quadrat" was deleted, because it had no microsite info.
- Duplicate entries of "*Nasella tenuis*" in "Species within quadrat" were deleted.

Australia
- Column names were changed to match the column names used by other collaborators.
- The Microsite variable was added, with 1 indicating bare microsites and 2 indicating nurse microsites.
- Replicate numbers in site 14 plot 3 were changed so that the replicate numbers followed on from that of the previous nurse.

Chile
- The 7th replicate of nurse *Mulinum spinosum* was missing from site plot 3 of site 25. From the data structure I deduced that the nurse's name was mislabelled as *Festuca pallescens*. Therefore, the 4th entry of *Festuca pallescens* in ID_Microsite of 25-3-Festuca pallescens-7 was renamed to *Mulinum spinosum*.

- The 16th replicate of nurse *Mulinum spinosum* was missing from site plot 3 of site 25. From the data structure I deduced that the nurse's name was mislabelled as *Festuca pallescens*. Therefore, the3rd, 4th and 5th entry of Festuca pallescens in ID_Microsite of 25-3-Festuca pallescens-16 was renamed to *Mulinum spinosum*.
- There was a second instance of 25-3-Mulinum spinosum-22 and an associated bare microsite. As it contains species also found in the first instance of 25-3-Mulinum spinosum-22, I assumed this was a numbering mistake. A replicate number of 26 was assigned to the 2nd and 3rd entry of 25-3-Mulinum spinosum-22 and to the 2nd entry of 25-3-Bare-22.
- Replicate numbers in plot 1, 2 and 3 of site 25 were changed so that the replicate numbers followed on from that of the previous nurse.
- Multiple quadrats were sometimes used to make up a bare microsite with an area equivalent to that of the nurse. In replicates where this was done, the average cover of species was calculated by dividing the sum of species cover by the number of quadrats used. After averaging, duplicate species names were deleted. Because of this, the Quadrat column becomes nonsensical (because the quadrat numbers do not follow on from each other) and should not be used in further analyses for this country dataset.

China: Chongfeng Bu
- Replicate numbers in plot 2 of site 29, and plot 2, 3 and 4 of site 30 were changed so that the replicate numbers followed on from that of the previous nurse.
- There was no bare microsite in replicate 12 in plot 1 of site 28. This replicate was deleted.

China: Xiaobing Zhou
- The first row was deleted, as it did not have any entries.
- Replicate numbers in plot 3 of site 34 were changed to follow on from the previous nurse.
- Duplicate entries of *Seriphidium borotalense* in replicates of plot 3 site 34 were deleted.

Iran: Mehdi Abedi
- The rows with "microsite_rep" = 41-1-Salsola arbusculiformis-1 and 41-2-Salsola arbusculiformis-1 were deleted. These rows had no entries in "Species within quadrat" and there were already other nurses in those replicates.
- Replicate numbers in plot 3 of site 41 and plot 1,2,3 of site 42 were changed so that the replicate numbers followed on from that of the previous nurse.

Iran: Mohammad Farzam
- The Bare Microsite of replicate 15 in plot 4 of site 45 with the nurse *Pistacia vera* has no entry in ID_Microsite. Assigned it a value of "Bare".
- Deleted other rows with no entry in "ID_Microsite", as they have no information in "Species within quadrat".
- Replicate numbers in plot 1,2,3,4 of site 43; plot 1,2,4 of site 44 and plot 2,3,4 of site 45 were changed so that the replicate numbers followed on from that of the previous nurse.
- Bare microsites were missing for replicates 2,4,6,7,9,10 with the nurse *Astragalus jolderensis* in plot 2 of site 43. Bare microsites without any entries in "Species within quadrat" were created for these replicates.
- Nurse microsites were missing for replicates 8 and 11 in plot 2 of site 43. Empty nurse microsites with ID_Microsite = *Astragalus jolderensis* and no entry in "Species within quadrat" were created.

- A bare microsite was missing for replicate 21 with the nurse *Acantholimon erinaceum* in plot 2 of site 43. A bare microsite without any entries in "Species within quadrat" was created for this replicate.
- A bare microsite was missing for replicate 4 with the nurse *Astragalus jolderensis* in plot 1 of site 44. A bare microsite without any entries in "Species within quadrat" was created for this replicate.
- A bare microsite was missing for replicate 4 with the nurse *Artemisia kopetdaghensis* in plot 2 of site 44. A bare microsite without any entries in "Species within quadrat" was created for this replicate.
- A bare microsite was missing for replicate 22 with the nurse *Ephedra major* in plot 2 of site 44. A bare microsite without any entries in "Species within quadrat" was created for this replicate.
- Bare microsites were missing for replicates 2,4,5,11,12,17,18,19,23,24 with the nurse *Pistacia vera* in plot 4 of site 45. Bare microsites without any entries in "Species within quadrat" were created for these replicates.
- Bare microsites were missing for replicates 1,2,3,4,5,6,7,8,9,10,11,13,15,17,18,19,20,22,24,25 with the nurse *Zygophyllum atriplicoides* in plot 4 of site 45. Bare microsites without any entries in "Species within quadrat" were created for these replicates.

Israel
- In site plot 3 of site 46 there were two nurse species in replicate 26. From the data structure it was deduced that the nurse *Sarcopoterium spinosum* belongs in the previous replicate. A new replicate number of 25 was assigned to the row with microsite_rep = "46-3-Sarcopoterium spinosum-26".

Namibia: Niels Blaum
- The "approx." and "comments" columns were removed.
- Where "Species within quadrat" = NA, "Cover" was changed to NA.
- Multiple quadrats were used in plots 196 and 197 to make up a bare microsite with an area equivalent to that of the nurse. In replicates where this was done, the average cover of species was calculated by dividing the sum of species cover by the number of quadrats used. After averaging, duplicate species names were deleted. Because of this, the Quadrat column becomes nonsensical and should not be used in further analyses.

Namibia: Lixing Wang
- Where "Species within quadrat" = NA, "Cover" was changed to NA.

South Africa
- The underscore between genus and specific epithets were removed so that species names are in the format "Genus species".

Spain: Fernando Maestre
- In plot 1,2,3 of site 84 the replicate numbers were changed to follow on from the preceding nurse.

Spain: Pedro Rey
- The column titled "Nd.individuals" was renamed to "Number.of.individuals".
- Column 20 contained no information, it was removed.
- The "Site" column, containing the names of plots was removed.

- The column titled "height..cm." was renamed to "height.cm.".
- In plot 1,2,4 of site 89; plot 1,2,3,4 of site 90 and plot 1,2,3 of site 91 the replicate numbers were changed to follow on from the preceding nurse.
- Rows in which the entry in "Species.within.quadrat" contained the word "muerto" (thus indicating a dead plant) were removed if it was not the only entry in the microsite. If the "muerto" entry was the only entry in a microsite, the entry in "Species.within.quadrat" was set to NA.