

Title: Effects of climate legacies on above- and below-ground community assembly

Running head: Climatic legacies drive community assembly

Manuel Delgado-Baquerizo^{1,2}, David J Eldridge³, Samantha K. Travers⁴, James Val⁵, Ian Oliver⁶, Andrew Bissett⁷.

1. Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309.
2. Departamento de Biología y Geología, Física y Química Inorgánica, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, Calle Tulipán Sin Número, Móstoles 28933, Spain.
3. Office of Environment and Heritage, c/- Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales, 2052, Australia.
4. Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales, 2052, Australia.
5. Office of Environment and Heritage, P.O. Box 363, Buronga, NSW, 2739, Australia.
6. Office of Environment and Heritage, P.O. Box 1477, Gosford, NSW, 2250, Australia.
7. CSIRO, Oceans and Atmosphere, Hobart, Tasmania, 7000, Australia

***Author for correspondence:**

Manuel Delgado-Baquerizo. Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309 E-mail: M.DelgadoBaquerizo@gmail.edu.au

Abstract

The role of climatic legacies in regulating community assembly of above- and below-ground species in terrestrial ecosystems remains largely unexplored and poorly understood. Here, we report on two separate regional and continental empirical studies, including >500 locations, aiming to identify the relative importance of climatic legacies (climatic anomaly over the last 20k years) compared to current climates in predicting the relative abundance of ecological clusters formed by species strongly co-occurring within two independent above- and below-ground networks. Climatic legacies explained a significant portion of the variation in the current community assembly of terrestrial ecosystems (up to 15.4%) that could not be accounted for by current climate, soil properties and management. Changes in the relative abundance of ecological clusters linked to climatic legacies (e.g., past temperature) showed the potential to indirectly alter other clusters, suggesting cascading effects. Our work illustrates the role of climatic legacies in regulating ecosystem community assembly and provides further insights into possible winner and loser community assemblies under global change scenarios.

Keywords: Paleoclimate, Bacteria, Fungi, Plants, Animals, Terrestrial ecosystems, Ecological networks.

Introduction

Current climate is known to be one of the major environmental filters shaping above and belowground community assemblies (Schleuning *et al.* 2016), as particular groups of species only occur under specific ranges of precipitation and/or temperature. Nevertheless, climatic conditions are dynamic, and have been shown to shift profoundly over millennia. Consequently, paleoclimatic filtering might have left a strong signature on the current above- (plants and animals) and below-ground (bacteria, fungi, protists and soil invertebrates) community assemblies found within ecological networks across entire terrestrial ecosystems. Although such an argument is intuitive conceptually, the relative importance of paloclimatic legacies (i.e., temperature and precipitation differences from the present to ~20k years ago; Fordham *et al.* 2017) compared with current climate filtering in predicting the assembly of entire ecological network of above- and below-ground communities has never been explicitly tested. Furthermore, no large scale studies have addressed this important research question. Here, we tested the hypothesis that historical climatic legacies (hereafter ‘climatic legacies’) explain important parts of the variation in ecosystem aboveground and belowground community patterns found within ecological networks that cannot be accounted for by current climates.

Studies over the past two decades provide strong evidence that climatic changes, since the last glaciation about 10k years ago (Fordham *et al.* 2017), are partly responsible for the current distribution of plants, animals and microbial communities in terrestrial ecosystems globally (Atkinson *et al.* 1987; Svenning J-C. *et al.* 2015; Lyons *et al.* 2016; Delgado-Baquerizo *et al.* 2017; Partel *et al.* 2017). Recent studies have also provided solid evidence that a knowledge of climatic legacies, can improve our predictions of the current distribution of specific groups of organisms including plants and microbes (Schleuning *et al.* 2016; Delgado-Baquerizo *et al.* 2017; Partel *et al.* 2017). Much less is known on the role of climatic legacies in driving ecological networks of above- and below-ground organisms. Plant and soil microbial communities comprise two components of the most important terrestrial food webs: aboveground and belowground. The first is essential for the provision of food and fibre and the second supports key soil processes such as litter decomposition and nutrient cycling, which in turn, supports plant productivity (Wardle *et al.* 2004; Hooper *et al.* 2000; de Vries *et al.* 2012). Because of their enormous functional importance, identifying new predictors that help explain the distribution of entire biotic community assemblies is one of the major endeavours in which scientists are immersed today. Moreover, future projections are conditional upon the past. Thus, a demonstrable link between climatic legacies and current

ecosystem community assemblies found within ecological networks would improve our capacity to predict how entire ecosystem community assemblies might respond to forecasted climate change, and the extent to which climatic changes might affect the myriad ecosystem services these communities provide.

Given the strong links between climatic legacies and the current distribution of particular groups of soil microbes and plants (Schleuning *et al.* 2016; Delgado-Baquerizo *et al.* 2017; Partel *et al.* 2017), we hypothesized that past climates might have triggered the current above- and belowground community assemblies in terrestrial ecosystems, i.e., the identity and abundance of coexisting multitrophic species within ecological networks that occur today. For example, locations with a positive anomaly in temperature or precipitation over the last 20k years might have resulted in a completely different biotic community assembly compared with locations with a negative anomaly or no change in temperature or precipitation, even if all these locations share the same current climate. If climatic legacies play a role in regulating the current network of ecological interactions, then climatic legacies might help to explain particular community assemblies that cannot be explained using only current climate data. This unexplained variation has hitherto generally been ascribed to stochasticity (e.g., Powell *et al.* 2015). Thus, climatic legacies might help us to explain why two locations with a similar current climate do not always lead to exactly the same community assembly.

We argue that Australia is one of the best locations on Earth to identify the role of climatic legacies in driving current ecosystem community assembly for three reasons. First, Australia has a long history of aboriginal occupation (> 60k years), characterised by a semi-sedentary, hunter-gatherer lifestyle (Hubble *et al.* 1983). Compared with other continents, Australia has a relatively recent history of European occupation (~ 200 years) and therefore a short history of intensive agriculture. Because of a short European history, more than 90% of Australia's land mass is still occupied by native vegetation and less than 6% is arable. Therefore, compared with other continents, the network of ecological interactions in Australian ecosystems is more likely to resemble those that existed prior to large-scale agricultural management. Second, contemporary agricultural land use in Australia is predominantly livestock grazing and cropping, and statistical models are able to account for the impacts of both land uses on our conclusions. Finally, given its continental scale, Australia experienced a wide range of climatic legacies over the past 20k years, including both positive and negative anomalies in temperature and precipitation variables (see

examples in Fig. S1). Consequently, Australia provides enough statistical variability to enable us to answer our primary research questions.

Herein, we used a combination of ecological network analyses and statistical modelling to evaluate the relative importance of climatic legacies compared to current climates in predicting the relative abundance of particular ecological clusters of strongly co-occurring species. Intuitively, we would expect these clusters to include species across multiple trophic levels and to be good surrogates of exclusive ecosystem community assemblies. We also aimed to identify the most important climatic legacies explaining the relative abundance of these ecological clusters and describe examples of specific species-species interactions within these clusters across different trophic groups (predator/prey) and associations (host/symbiont). To address our research questions, we used two independent datasets from Australia, which included >500 locations at both regional and continental scales. The first dataset, which included information, from 108 “natural” locations in eastern Australia on the composition of plant and animal species, was used to build an aboveground (plants and animals) correlation network. The second dataset, including 375 “natural” and 60 cultivated locations across mainland Australia, contained information on the composition of soil bacteria and eukaryotes (fungi, protists and soil invertebrates). This dataset was used to build a belowground correlation network.

Materials and Methods

Aboveground network

Our aboveground network study was conducted at 108 sites across a large area (> 500 km²) of eastern Australia (Fig. S1). This survey was undertaken in three semi-natural woodland communities dominated by blackbox (*Eucalyptus largiflorens*), white cypress pine (*Callitris glaucophylla*) and river red gum (*Eucalyptus camaldulensis*). These three communities include sites used extensively for livestock grazing, large areas dedicated to conservation (national parks, nature reserves) and smaller areas devoted to native forestry, but excluded any areas that were cultivated or supported crops. In these locations, we undertook multiple vegetation and animal surveys targeting grasses, forbs, woody plants, birds, mammals, reptiles, amphibians and invertebrates (see Appendix S1 for sampling details).

Belowground network

We used a subset of sample locations from the Biome of Australia Soil Environments (BASE) project (Fig. S1) for our belowground network (soil bacteria, fungi, protists and soil invertebrates). This subset includes data on the composition of bacterial, fungal and eukaryotic communities across 439 locations belonging to “natural” (379) and agricultural

(60) (wheat and cotton crops) ecosystems from Australia. Samples were collected between 2011 and 2014. In each location, a 25 x 25m plot was established. Soil samples (top 10cm) were collected according to the methods described in Bissett *et al.* (2016). The community composition of soil bacteria, fungi, protists and soil invertebrates was determined using amplicon sequencing with the Illumina Miseq platform (see Appendix S2 for details).

Climate data

For all sites surveyed, we obtained six climatic variables for current climate and climate in the Last Glacial Maximum from the Worldclim database (www.worldclim.org) (Hijmans *et al.* 2005). These variables include mean precipitation (MAP), maximum and minimum temperature (MAXT and MINT), mean annual precipitation and temperature seasonality (PSEA and TSEA) and mean diurnal temperature range (MDR). We selected these six variables as they provide a good approximation of the quantity and variability of precipitation and temperature. In addition, these six variables did not suffer from strong multi-collinearity (Pearson's $r < 0.8$; Katz 2006). In the case of Last Glacial Maximum climate, we used estimates provided by the Community Climate System Model (CCSM4; www.worldclim.org) (Bystrakova *et al.* 2013; Tallavaara *et al.* 2015). We used data at a 2.5 minutes resolution (~4.5km at Equator), as this is the highest resolution available for the Last Glacial Maximum period. Previous studies have demonstrated that the Last Glacial Maximum information used here, largely resemble information coming from other climatic models (Delgado-Baquerizo *et al.* 2016a) and spatial and temporal resolutions (Delgado-Baquerizo *et al.* 2017).

Climatic legacies

Climatic legacies were calculated as the differences between an estimate of six climatic variables (amount and variability in precipitation and seasonality) 20k ybp and another estimate for these variables at the present day (Fordham *et al.* 2017) as shown in Delgado-Baquerizo *et al.* (2017). In particular, the climatic legacy for each climatic variable is calculated as the mathematical difference in the values for each climatic variable from Last Glacial Maximum and current climates (e.g., $\text{Annual precipitation}_{\text{Current climate}} - \text{Annual precipitation}_{\text{Last Glacial Maximum}}$) for each site. This difference provides us with a measure of climatic legacies; increases, declines or a lack of change in a particular climatic condition with time- in each of the sites surveyed from the different datasets. A recent cross-validation of the climatic legacy indexes used here is given in Appendices 1-3 in Delgado-Baquerizo *et al.* (2017) and Appendix S1 in Delgado-Baquerizo *et al.* (2016a). Note that the climatic legacy index used here is based on the differences between two single snapshots in time

(Current vs. Last Glacial Maximum climates), thus calculation of climate legacy comes with a number of inherent and important assumptions (Fordham *et al.* 2017). For example, although we assume that change in precipitation and temperature gradually occurred with time during last 21k years, we would like to acknowledge that most abrupt changes in climate occurred prior to 10k YBP (see Fordham *et al.* 2017). Even so, our climatic legacy index still allowed us to address our research question of whether the signature of climatic legacies on the network of interactions of aboveground and belowground can still be detected today. Further discussion on this point is available at Appendices 1-3 in Delgado-Baquerizo *et al.* (2017).

Soil properties and current management

Soil properties including texture (% sand content), pH, soil C and P were available from the two datasets used here (Appendix S3). Current management including intensity of grazing by cattle and the incidence of cropping (only applicable to the belowground dataset) was included in our statistical models (see below) to account for impacts from recent management in the network of interactions of aboveground and belowground. We used cattle density as our proxy of current management as grazing by cattle is one of the major drivers of grazing-induced degradation in Australia over the past 200 years (see Appendix S3 for details).

Network analyses

Network analyses were conducted separately for the aboveground and belowground network Australian datasets. In both cases, we identified ecological clusters of strongly associated taxa using correlation networks ('co-occurrence network') and the following protocol. Our aboveground network contained 1280 nodes (species of vascular plants, mammals, birds, reptiles, amphibians, ants, beetles, centipedes, cockroaches, crickets, scorpions and spiders). In the case of the belowground networks, our datasets included 95,208 Operational Taxonomic Units (OTUs) of bacteria, fungi, protists and soil invertebrates. These OTUs (aka phylotypes) were calculated at 97% sequence similarity and can be considered to be analogous to "species". However, because of the large number of microbial 'species' compared with other groups (plants and animals) and the need to restrict analyses to a manageable network of interactions, we focused on the dominant microbial OTUs (top 10% species sorted by dominance, as described in Soliveres *et al.* 2016). Dominant species for bacteria, fungi and other eukaryotes were obtained independently for these organisms from their original OTU tables. These bacterial, fungal and other eukaryotic taxa were then merged

into a single abundance table. This resulted in a dataset with 9502 taxa including 4953 bacteria (~80% of all bacterial phylotypes), 2321 fungi (~80% of all fungal phylotypes) and 2228 other eukaryotes phylotypes (~80% of all eukaryotic phylotypes). We then calculated all pairwise Spearman's rank correlations (ρ) between all soil plant/animal and soil microbe/animal taxa. We focused exclusively on positive correlations as they provide information on microbial taxa that may respond similarly to environmental conditions (Barberan *et al.* 2012). We considered a co-occurrence to be robust if the Spearman's correlation coefficient ρ was > 0.50 and $P < 0.01$ (see Barberan *et al.* 2012 for a similar approach). The network was visualized with the interactive platform Gephi (Bastian *et al.* 2009). Finally, we used default parameters from the interactive platform Gephi to identify ecological clusters (aka modules) of soil taxa strongly interacting with each other (Bastian *et al.* 2009). We then computed the relative abundance of each ecological cluster by averaging the standardized relative abundances (z-score) of the taxa that belong to each ecological cluster. By standardizing our data, we ruled out any effect of merging data from different soil groups: plants/animals and soil microbes/animals. In addition, we also used an alternative approach and calculated the relative abundance of ecological clusters after centered log-ratio transformation. Information on functional traits for fungal taxa within each ecological cluster (which is unavailable for bacteria), was obtained from the online application FUNGuild described in Nguyen *et al.* (2016).

Variation partitioning modelling

We used Variation Partitioning (Legendre *et al.* 2008) to quantify the relative importance of four groups of predictors: 1) six climatic legacies, 2) six climatic variables from current climate, 3) current management (cattle density in regional Australia and cattle density and cropping in continental Australia) and 4) soil properties (pH, % of sand, soil C and P) as predictors of the relative abundance of ecological clusters in the (1) Aboveground network and (2) Belowground network. This analysis allowed us to identify whether climatic legacies can explain a unique portion of the variance that is not explained by current climate or management (Legendre *et al.* 2008). Note that adjusted coefficients of determination (R^2) in multiple regression and canonical analysis can, on occasion, take negative values (Legendre *et al.* 2008). Negative values in the variance explained for a group of predictors on a group of response variable are interpreted as zeros, and correspond to cases in which the explanatory variables explain less variation than that explained using random normal variables (Legendre

et al. 2008). In all cases, Variation Partitioning analyses were conducted with the R package Vegan (Oksanen *et al.* 2015).

Random Forest modelling

We conducted a classification Random Forest analysis (Breiman 2001) as described in Delgado-Baquerizo *et al.* (2016b) to identify the major predictors of the relative abundance of ecological clusters in the two networks. Our list of predictors included six climatic legacies, six climatic variables from current climate, soil properties (pH, % of sand, soil C and P) and current management (cattle density and/or cropping). These analyses were conducted using the rfPermute package (Archer *et al.* 2016) of the R statistical software (<http://cran.r-project.org/>). We also repeated these analyses using an alternative Random Forest approach using the gradientforest R package (Strobl *et al.* 2008; Ellis *et al.* 2012).

Structural equation modeling

We used structural equation modeling (SEM) (Grace 2006) to evaluate effects of climatic legacies (i.e., temperature and precipitation differences between estimated climate about 20k ybp and current climatic estimates) on the relative abundance of ecological clusters in the two networks, after accounting for spatial autocorrelation (latitude and longitude), soil properties (pH, % of sand, soil C and P), current management (cattle density and/or cropping) and current climate. Our *a priori* model is shown in Fig. S3. The use of SEM is particularly useful in large scale correlative studies, as it allows the partitioning of causal influences among multiple variables, and separation of the direct and indirect effects of model predictors (Grace 2006). We then tested the goodness of fit of our models. The goodness of fit of SEM models was checked following Schermelleh-Engel *et al.* (2003). There is no single universally accepted test of overall goodness of fit for SEM, applicable in all situations regardless of sample size or data distribution (Schermelleh-Engel *et al.* 2003). We used the χ^2 test (χ^2 ; the model has a good fit when $0 \leq \chi^2/DF \leq 2$ and $0.05 < P \leq 1.00$) and the root mean square error of approximation (RMSEA; the model has a good fit when $0 \leq RMSEA \leq 0.05$ and $0.10 < P \leq 1.00$) (Schermelleh-Engel *et al.* 2003). Our *a priori* models attained an acceptable/good fit by all criteria in all cases, and thus no post hoc alterations were made. With a good model fit, we were free to interpret the path coefficients of the model and their associated *P* values. SEM models were conducted with the software AMOS 20 (IBM SPSS Inc, Chicago, IL, USA).

Results

We first generated two separate ecological networks for 1) aboveground and 2) belowground using information from the two independent datasets. Using the approach described in the Methods section, we identified and calculated the relative abundance of six and seven major ecological clusters of species co-occurrence for aboveground and belowground networks, respectively (Fig. 1). All taxa included within each ecological cluster for the two networks, and additional functional information on these taxa, are shown in Figs. S4 and S5 and Table S1. These ecological clusters include multiple species linked by potential ecological interactions such as predator/prey, host/parasite, host/symbiont, as well as different trophic levels, e.g., primary producers and primary consumers (Table S1; Figs S4 and S5). We found a highly significant correlation between the relative abundance of the ecological clusters calculated as explained above and the same clusters calculated after using the centered log-ratio transformation ($\rho > 0.90$; $P < 0.001$; Table S2).

Our variation partitioning model suggested that climatic legacies explained a unique portion of the variation for particular ecological clusters that could not be accounted for by measures of current management, soil properties or current climates (Fig. 1). Climatic legacies explained a unique and significant portion of the variation of ecological clusters in five out of six ecological clusters for our aboveground network (AG#) and for seven out of seven ecological clusters for our belowground network (BG#) (Fig. 1; Table S3). This was especially noticeable for AG#1 (6 out of 58% of variation explained) and for BG#1 (7.5 out of 68% of variation explained) and BG#3 (15.5 out of 47% of variation explained). As expected, current climate and soil properties, routinely proposed as the dominant drivers of ecosystem community assembly at large spatial scales, also explained a unique portion of the variation in all ecological clusters (Fig. 1; Table S3). Management was also important for some clusters (BG#0, 1, 5 and 6), but not for others (Fig. 1; Table S3).

Our Random Forest analyses suggested that climatic legacies were as important as, or more important than, current climate in predicting the relative abundance of ecological clusters within our two networks, and after accounting for soil properties and management (Figs. S6-S9). Temperature legacies, maximum temperature (MAXT) and temperature seasonality (TSEA), were more consistent than precipitation legacies in predicting the relative abundance of ecological clusters within the aboveground and belowground networks (Figs. S6-S9). However, mean precipitation (MAP) and precipitation seasonality (PSEA) were also reported to be a key climatic legacy predicting the relative abundance of ecological assemblies in both ecological networks (Figs. S6-S9). These results suggest that climatic legacies have left a detectable signature on the contemporary ecosystem community assembly

of unique clusters of plant, animal and microbial species strongly co-occurring with each other (Table S1; Figs S4 and S5). Importantly, we found a statistically significant correlation between the Random Forest importances across predictors calculated from each ecological cluster using the rfPermute and gradientforest R packages (Table S4).

We used SEM, to further clarify the role of climatic legacies in predicting the relative abundance of contemporary ecosystem community assemblies, independently, for our two datasets. Although this is quite a conservative procedure, we still found multiple direct effects of climatic legacies, from all climatic variables studied, on the relative abundance of particular ecological clusters in both networks (Fig. 2). Remarkably, increases in maximum temperature legacies had a direct positive effect on the relative abundance of AG#1 and BG#1. In other words, locations with a positive anomaly for maximum temperature over the past ~20k years might have promoted the relative abundance of species within AG#1 and BG#1 (Fig. 3). Other highly significant climatic legacy effects also included a direct negative effect of temperature seasonality on the relative abundance of AG#2, and direct positive effects of diurnal temperature range (MDR) and TSEA on BG#2 and 0. Note that for simplicity, Figure 2 only included direct effects with a $P < 0.01$ (see Table S5 for direct effects with a < 0.01 $P < 0.05$). Remarkably, increases in a given ecological cluster were often followed by declines in the relative abundance of other ecological clusters, as supported by the multiple indirect effects among the relative abundance of ecological clusters in our aboveground and microbe-animal networks (e.g., AG/BG#2 vs. AG/BG#3)(Fig. 2 and 3). Moreover, we also detected multiple indirect effects of climatic legacies on the relative abundance of ecological clusters *via* changes in soil properties for the two studied (Fig. 2).

Discussion

Our findings provide novel evidence that past climates likely played an important role in regulating the relative abundance of the major ecosystem community assemblages that we observe today, helping to explain a unique portion of the variation in the distribution of particular communities that has generally been attributed to stochasticity (e.g. Powell *et al.* 2015). More specifically, climatic legacies might regulate the relative abundance of multiple ecological clusters formed by species strongly co-occurring within two independent above- and below-ground networks (shown in Table S1). These co-occurring taxa included multiple potential ecological interactions such as predator/prey, host/parasite, host/symbiont, as well as different trophic levels, e.g. primary producers and primary consumers (Table 1; Figs S3 and S4). For example, AG#1 contains multiple potential predator/prey interactions including (1) those of the birds species *Struthidea cinerea* and *Turnix velox*, with multiple potential

plant and arthropod preys (Table 1; Table S1), (2) those from the barking spider (*Selenocosmia stirlingi*) and the scorpion (*Lychas jonesae*) with the abundance of potential ant, beetle and cricket prey or (3) those from the lace monitor (*Varanus varius*) and the skink (*Ctenotus leonhardii*), both of which are related to the abundance of a wide range of arthropod species that they prey on. Similarly, BG#1 contains potential predator/prey relationships such as the reported high correlation between the protozoan *Cercozoa* and *Ciliophora* and their common prey soil bacteria. This ecological cluster also contains multiple fungal plant pathogens such as *Truncatella*, *Coniothyrium* and *Phoma* sp., with implications for plant communities co-existing with our belowground network. The relative abundance of all species and potential interactions within AG#1 and BG#1 might have been promoted by positive anomalies in maximum temperature (Figs. 2 and 3). Supporting this result, BG#1 contains multiple phylotypes of bacteria that have previously been reported to respond positively to increases in temperature including those from the genera *Candidatus*, *Koribacter*, *Bacillus*, *Burkholderia* and *Rhodoplanes* (Oliverio *et al.* 2016). In other words, locations with the highest positive anomalies in maximum temperature might now support a greater abundance of species within BG#1 than locations with negative anomalies or no changes in maximum temperature over the past 20k years.

Interestingly, increases in a given ecological cluster were often followed by declines in the relative abundance of other ecological clusters, as supported by the multiple indirect effects among the relative abundance of ecological clusters in our aboveground and microbe-animal networks. Such results suggest that increases in the relative abundance of particular ecological clusters resulting from temperature legacies might have had multiple cascading effects on other ecological clusters. Thus, the negative relationship between BG/AG#1 with BG/AG#2, might also lead to cascading effects on the relative abundance of BG/AG#3 clusters, which were negatively related to BG/AG#2 clusters in both networks (Figs. 2 and 3). Thus, climatic legacies might also have multiple indirect negative or positive effects on the relative abundance of the ecological clusters within our two networks, as supported by our structural equation models (Fig. 2). Cluster BG#2 in the microbe-animal network, contained multiple probable mycorrhizal species such as *Entoloma*, *Glomus* and *Claroideoglomus*, which might have positive effects on plant species linked to this soil microbial-network; some identified using molecular techniques (Table S1). This ecological cluster also includes potential predator/prey relationships between soil amoeba, and ciliates, with bacteria and plant pathogens (e.g. *Gibberella intricans*). Moreover, for the aboveground network, AG#2 was characterized by the potential producer/consumer link between emus (*Dromaius*

novaehollandiae) and the fruits of *Lycium ferocissimum*, *Eremophila debilis* and *Einadia* spp. (Noble 1991). Our findings suggest that the relative abundance and potential interactions among species within all of these ecological clusters are highly sensitive to anomalies in maximum temperatures and their cascading effects (Fig. 3). Changes in ecological clusters #2 in both networks, linked to climatic anomalies, might in turn have multiple cascading effects on the relative abundance of ecological clusters #3 in both networks. BG#2 is characterized by potential parasite/host interactions between *Gregarina* sp. and soil arthropods (Omoto and Cartwright 2003), predator/prey interactions between phylotypes from phylum *Cercozoa* (protist) and bacteria (Table S1) and plant-fungal interactions of symbiosis (e.g. *Auritella* sp.) and pathogenesis (e.g. *Devriesia* sp.). Similarly, AG#3 is characterized by potential predator/prey interactions among plants, insects, lizards and amphibians.

Although the effects of current or climatic legacies on the community assembly of terrestrial ecosystems are not directly comparable to those from on-going changes in climate, our network approach still has the potential to provide insights into the role of climate change in predicting possible winner and loser community assemblies in response to climate change. Specifically, our study provides a compendium of species from particular ecological clusters that are expected to be highly sensitive to changes in climatic conditions. For example, current maximum temperature, one of the major climatic legacies, is also positively and strongly influencing the abundance of BG#1 (Fig. 2) and AG#1 (SEM direct effect = 1.45; $P = 0.044$; Table S3). This suggests that further increases in temperature predicted by the end of this century might continue to promote the relative abundance of species and interactions within this ecological cluster, largely to the detriment of those in BG#2 and AG# 2, with potential cascading effects on other ecological clusters. In fact, our findings suggest that positive anomalies of maximum temperature of up to 4°C –comparable to those predicted for climate change already had a massive effect on the relative abundance of particular ecological clusters (Fig. 3). Predicted impacts of changes in precipitation with climate change (Huang *et al.* 2016) could also be inferred from our network approach. For instance, for our aboveground network, current precipitation seasonality, whose legacy was positively related to the abundance of AG#4, indirectly via changes in soil pH, is still having an effect on the abundance of this ecological cluster, characterized for the potential interactions between the sand monitor (*Varanus gouldii*) and its prey items lycosid spiders (*Lycosid* spp.) and scorpions (*Lychas* spp.). Other examples, of key climatic legacies that still drive the relative abundance of particular ecological clusters can be inferred from Fig. 2 and Tables S1 and S3. These climatic changes could also alter the proportion of native species or the proportion of

taxa coming from different species. For example, BG#1 in our microbe-animal network comprised mainly of phylotypes of strongly occurring bacteria and fungi, but BG# 2 also contained multiple soil animal taxa.

Finally, as expected (e.g. Gossner *et al.* 2016), current management also influenced the relative abundance of ecosystem community assemblies. For example, for our belowground network, we found strong direct effects of cropping and cattle density on the relative abundance of the multiple ecological clusters within this network (Fig. 2). Of special interest is the negative effect of cropping on the relative abundance of BG#1 in our microbe-animal network, which might potentially reverse part of the climatic legacies from maximum temperature (explained above) on this ecological cluster. Interestingly, BG#2 in the microbial-animal network, which was indirectly negatively affected by the maximum temperature legacy, seems to benefit from cropping and cattle density impacts (Fig. 2), though indirectly, potentially helping to reverse climatic legacies on BG#1. All of these results accord with previous studies suggesting that human activities can erase part of the climatic legacies of temperature and precipitation on the current distribution of soil organisms (Delgado-Baquerizo *et al.* 2017). However, in general, management measured as cattle density did not influence the relative abundance of aboveground clusters as supported by our Variation Partitioning, Random Forest and Structural Equation Modeling analyses. The only ecological cluster affected by cattle density in this network was AG#5, shown in our Random Forest results (Fig. S6).

Together, our work suggests that climatic legacies have left a statistically significant signature on the contemporary below- and above-ground community assemblies and can now explain a unique portion of the distribution in particular ecological clusters from terrestrial ecosystems. This is true even after accounting for key predictors such as location, soil properties, current climate or management, all of which are routinely proposed as drivers of ecosystem community assemblies at large spatial scales. These findings also advance our understanding of the links between particular climatic legacies and the relative abundance of species and potential interaction within ecological clusters across a broad range of ecosystem types at the continental scale. Moreover, we found that climatic anomalies might have led to multiple cascading effect on the relative abundance of ecological clusters in terrestrial ecosystems. We also found that current management influences can potentially reverse part of the impacts of climatic legacies, which occurred during the last 20k years, on particular ecological clusters. Such knowledge can potentially help us to better understand changes in

particular ecosystem community assemblies in response to on-going global environmental change including land use intensification and climate change, with important implications for future sustainable management and conservation policies.

Acknowledgments

M.D-B. acknowledge support from the Marie Skłodowska-Curie Actions of the Horizon 2020 Framework Programme H2020-MSCA-IF-2016 under REA grant agreement n° 702057. D.J.E. was supported by the Hermon Slade Foundation. We would like to acknowledge the contribution of the Biomes of Australian Soil Environments (BASE) consortium (<https://data.bioplatforms.com/organization/pages/bpa-base/acknowledgements>) in the generation of data used in this publication. The BASE project is supported by funding from Bioplatforms Australia through the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS). We also thank the New South Wales Office of Environment and Heritage for the financial support responsible for the aboveground dataset.

Statement of authorship: M.D-B. conceived this study. The aboveground dataset was compiled by D.J.E., S.K.T., J.V. and I.O. The belowground dataset was compiled by A.B. M.D-B. conducted network and statistical modelling. The manuscript was written by M.D-B, edited by D.J.E., and all co-authors significantly contributed to improve it.

Data accessibility:

Data associated with this paper has been deposited in figshare: <https://figshare.com/s/899e610ba9fd353cae49> (10.6084/m9.figshare.6217154). The raw sequence data used in this study is available at <https://data.bioplatforms.com/organization/about/australian-microbiome>.

References

- Archer, E. (2016). rfPermute, Estimate Permutation p-Values for Random Forest Importance Metrics. R package version 1.5.2.
- Atkinson, T.C., Briffa, K.R. & Coope, G.R. (1987). Seasonal temperatures in Britain during the past 22,000 years, reconstructed using beetle remains. *Nature* **325**, 587-592.
- Barberán, A., Bates, S.T., Casamayor, E.O. & Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343-351.
- Bastian M., Heymann S. & Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

500 Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P.M., Reith, F., Dennis, P.G. *et al.* (2016).
501 Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity
502 database *GigaSci*, 20165: 21.

503 Breiman. L. (2001). *Machine Learning*, **45**, 5.

504 Bystriakova, N., Ansell, S.W., Russell, S.J., Grundmann, M., Vogel, J.C. & Schneider, H.
505 (2014). Present, past and future of the European rock fern *Asplenium fontanum*:
506 combining distribution modelling and population genetics to study the effect of climate
507 change on geographic range and genetic diversity. *Ann. Bot.* **113**, 453–465

508 De Vries, F.T., Liiri, M., Bjørnlund, L., Bowker, M., Christensen, S., Setälä, H.M. *et al.*
509 (2012). Land use alters the resistance and resilience of soil food webs to drought. *Nat.*
510 *Clim. Change* **2**, 276–280.

511 Delgado-Baquerizo, M., Reich, P.B., García-Palacios, P., Milla, R. (2016a) Biogeographic
512 bases for a shift in crop C:N:P stoichiometries during domestication. *Ecol. Lett.* **19**, 564-
513 575.

514 Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Trivedi, P., Osanai, Y., Liu, Y-R. *et al.*
515 (2016b) Carbon content and climate variability drive global soil bacterial diversity
516 patterns *Ecol. Monogr.* **86**, 373-380.

517 Delgado-Baquerizo, M., Bissett, A., Eldridge, D.J., Maestre, F.T., He, J-Z., Wang, J-T. *et al.*
518 (2017). Paleoclimate explains a unique proportion of the global variation in soil bacterial
519 communities. *Nat. Ecol. Evol.* **1**, 1339–1347.

520 Ellis, N., Smith, S.J., and Pitcher, C.R. (2012) Gradient Forests: calculating importance
521 gradients on physical predictors. *Ecology* **93**, 156–16.

522 Fordham, D.A., Saltré, F., Haythorne, S., Wigley, T.M.L., Otto-Bliesner, B.L., Chan, K. *et al.*
523 (2017). PaleoView: a tool for generating continuous climate projections spanning the last
524 21000 years at regional and global scales. *Ecography* **40**, 1448-1458.

525 Gossner, M.M., Lewinsohn, T.M., Kahl, T., Grassein, F., Boch, S., Prati, D. *et al.* (2016).
526 Land-use intensification causes multitrophic homogenization of grassland communities.
527 *Nature* **540**, 266-269.

528 Grace, J.B. (2006). *Structural Equation Modeling Natural Systems* (Cambridge Univ. Press,
529 Cambridge).

530 Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, Jarvis, A. (2005). Very high resolution
531 interpolated climate surfaces for global land areas. *International Journal of Climatology*
532 **25**, 1965-1978.

533 Hooper, D.U., Bignell, D.E., Brown, V.K., Brussaard, J.M., Dangerfield, J.M., Wall, D.H. *et*
 534 *al.* (2000). Interactions between aboveground and belowground biodiversity in terrestrial
 535 ecosystems: patterns, mechanisms, and feedbacks. *BioScience* **50**, 1049-1061.
 536 Huang, J., Yu, H., Guan, X., Wang, G. & Guo, R. (2016) Accelerated dryland expansion
 537 under climate change. *Nat. Clim. Change* **6**, 166–171.
 538 Hubble, G.D., Isbell, R.F. & Northcote, K.H. (1983). Features of Australian soils. In: Soils,
 539 an Australian Viewpoint (Divn. Soils CSIRO Australia), pp. 17-47. CSIRO/Academic
 540 Press.
 541 Katz, M.H. (2006). Multivariable Analysis: A Practical Guide for Clinicians and Public
 542 Health Researchers (Cambridge University Press, Cambridge, UK, 2006).
 543 Legendre, P. (2008). Studying beta diversity: ecological variation partitioning by multiple
 544 regression and canonical analysis. *J. Plant Ecol.* **1**, 3-8.
 545 Lyons, S.K., Amatangelo, K.L., Behrensmeyer, A.K., Bercovici, A., Blois, J.L. David, M. *et*
 546 *al.* (2016). Holocene shifts in the assembly of plant and animal communities implicate
 547 human impacts. *Nature* **529**, 80–83.
 548 Nguyen, N.H., Song, Z., Bates, S.T., Branco, S., Tedersoo, L., Menke, J. *et al.* (2016).
 549 FUNGuild: An open annotation tool for parsing fungal community datasets by
 550 ecological guild. *Fungal Ecol.* **20**, 241-248.
 551 Noble, J.C. (1991). On ratites and their interactions with plants. *Revista Chilena de Historia*
 552 *Natural* **64**, 85-118.
 553 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. *et al.* (2015).
 554 Vegan: Community Ecology Package. R package version 2.3-0.
 555 Oliverio, A., Bradford, M.A. & Fierer, N. (2016). Identifying the microbial taxa that
 556 consistently respond to soil warming across time and space. *Glob. Change Biol.* **23**,
 557 2117–2129.
 558 Omoto, C.K. & Cartwright, D.C. (2003). Investigating the diversity of parasitic protozoa
 559 using gregarine parasites of invertebrates, pp. 77-85. In: Tested Studies for Laboratory
 560 Teaching, (M.A. O'Donnell, Ed.). Proc. 24th Workshop/Conference of the Association
 561 for Biology Laboratory Education (ABLE), pp. 334.
 562 Partel, M., Öpik, M., Tedersoo, L., Szava-Kovats, R., Rosendahl, S., Rillig, M.C. *et al.*
 563 (2017). Historical biome distribution and recent human disturbance shape the diversity
 564 of arbuscular mycorrhizal fungi. *New Phytol.* **216**, 227-238.

- Powell, J.R., Karunaratne, S., Campbell, C.D., Yao, H., Robinson, L. & Singh, B.K. (2015). Deterministic processes vary during community assembly for ecologically dissimilar taxa. *Nat. Commun.* **6**, 8444.
- Schermelleh-Engel, K., Moosbrugger, H. & Muller, H. (2003). Evaluating the fit of structural equation models, tests of significance descriptive goodness-of-fit measures. *Methods Psychol. Res. Online* **8**, 23–74.
- Schleuning, M., Fründ, J., Schweiger, O., Welk, E., Albrecht, J., Albrecht, M. *et al.* (2016). Ecological networks are more sensitive to plant than to animal extinction under climate change. *Nat. Comm.* **7**, 13965
- Soliveres, S., Manning, P., Prati, D., Gossner, M.M., Alt, F., Arndt, H. *et al.* (2016). Locally rare species influence grassland ecosystem multifunctionality. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150269.
- Strobl, C. Boulesteix, A.-L., Kneib, T., Augustin, T. and Zeileis, A. (2008) Conditional variable importance for random forests. *BMC Bioinformatics* **9**, 307–317.
- Svenning J-C., Eiserhardy, W.L., Normand, S., Ordenez, A. & Sandel, B. (2015). The influence of paleoclimate on present-day patterns in biodiversity and ecosystems. *Ann. Rev. Ecol. Evol. Syst.* **46**, 551-572.
- Tallavaaraa, M., Luoti, M., Korhonen, N., Järvinen, H. & Seppä, H. (2015). Human population dynamics in Europe over the Last Glacial Maximum. *Proc. Natl. Acad. Sci. USA* **112**, 8232–8237.
- Wardle, D.A., Bardgett, R.D., Klironomos, J.N., Setälä, H., van der Putten, W.H. & Wall, D.H. (2004). Ecological linkages between aboveground and belowground biota. *Science* **304**, 1629-1633.

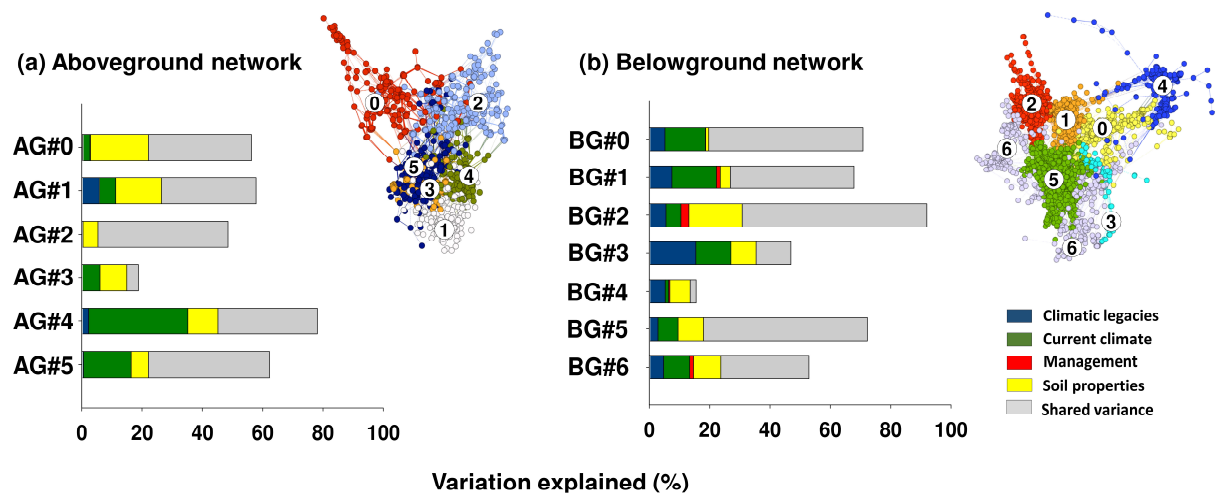
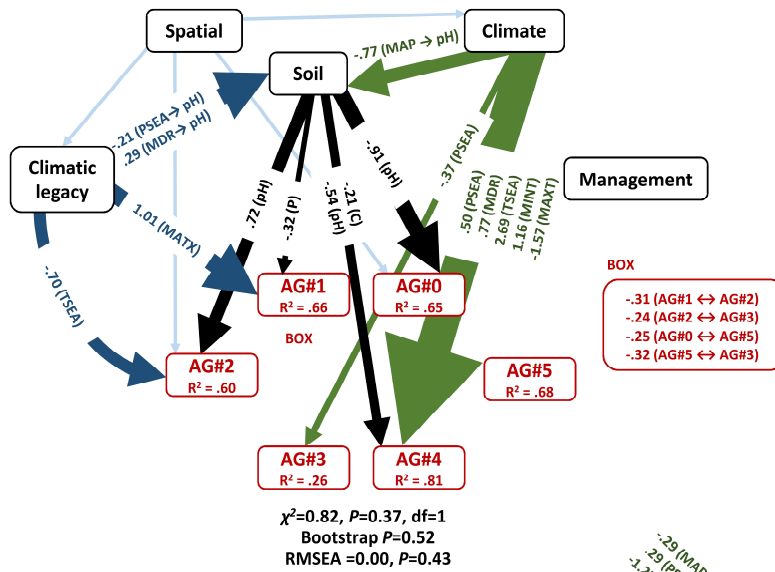


Figure 1. Relative contribution of the different predictors used to model the relative abundance of ecological clusters within our aboveground and belowground networks. Upper right panels represent network diagrams with nodes coloured by each ecological cluster within our aboveground and microbe-animal networks. A characterization of the taxa within each ecological cluster is available in Table S1. Bottom left panels represent results from Variation Partitioning modelling aiming to identify the percentage variance of relative abundance of ecological clusters explained by climatic legacies, current climate and management. Associated P-values to the relative contribution of the different predictors are available in Table S3. AG = Aboveground network. BG = Belowground network.

(a) Aboveground network



(b) Belowground network

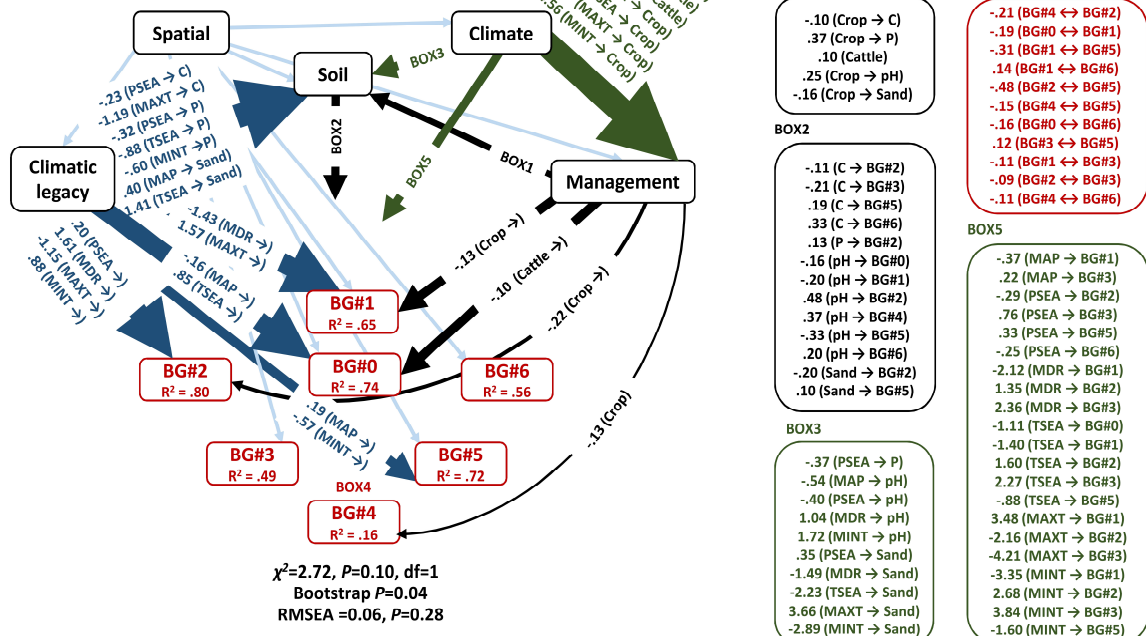


Figure 2. Mechanistic modeling identifying the direct and indirect effects of climatic legacies on the relative abundance of ecological clusters within our aboveground and belowground networks. For simplicity, only effects with a $P < 0.01$ are reported here. The rest of significant effects are available in Table S5 ($0.01 < P < 0.05$). Numbers adjacent to arrows indicate the effect-size. R^2 denotes the proportion of variance explained. The size of the arrow is proportional to the effect size (but in the case of spatial influence). Climatic legacies, current climate and management predictors are included in our models as independent observable variables, however we grouped them in the same box in the model for graphical simplicity. AG = Aboveground network. BG = Belowground network.

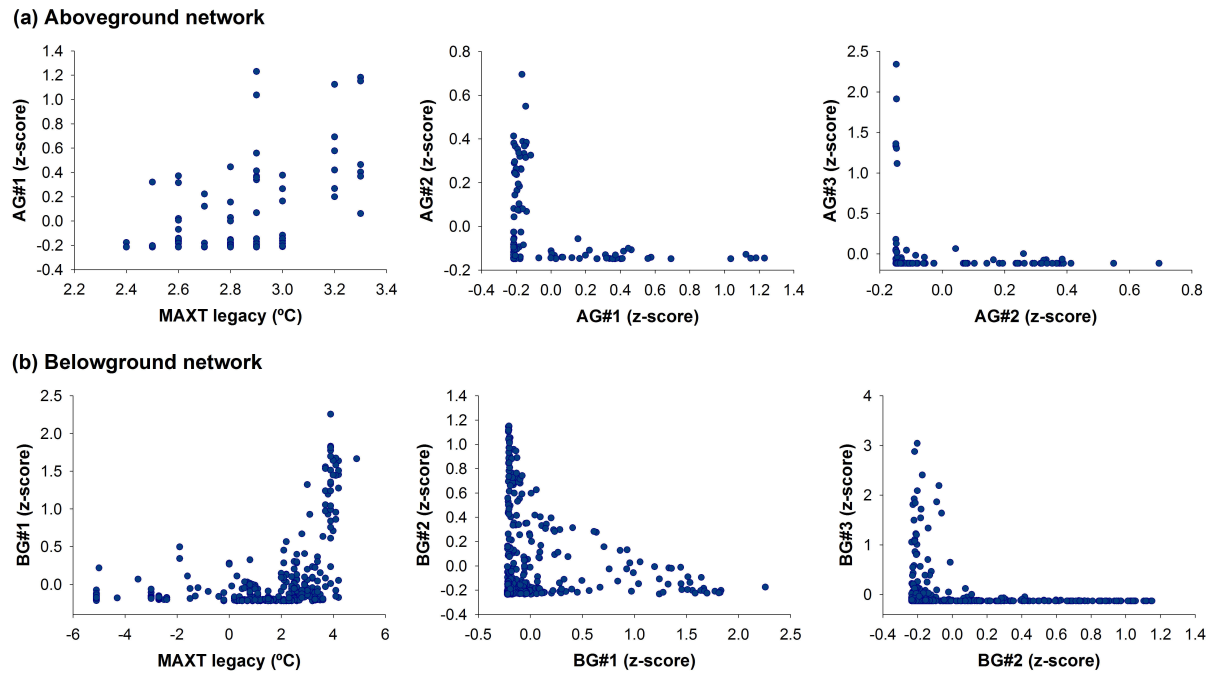


Figure 3. Selected relationships from our SEMs. Panels include relationship between maximum temperature legacy and ecological clusters #1 within our aboveground and belowground networks. Also, selected relationships between ecological clusters #1 and 2, and ecological clusters #2 and 3 in both independent networks. AG = Aboveground network. BG = Belowground network. See Fig. S10 for an alternative version of this figure using ecological clusters calculated after centred log-ratio transformation and showing similar results.

Supplementary Information

Effects of climate legacies on above- and below-ground community assembly

Manuel Delgado-Baquerizo, David J Eldridge, Samantha K. Travers, James Val, Ian Oliver, Andrew Bissett.

Correspondence to: M.delgadobaquerizo@gmail.com

This PDF file includes:

Appendices S1-S3

Figures S1-S7

Tables S1-S3

Appendix S1. Information on the plant and animal sampling to obtain the data used to build our plant-animal network (regional scale).

Each site comprised a 200 m long transect running perpendicular to the nearest livestock watering point, which was generally an earthen dam. Along this transect we positioned five 25 m² (5 m x 5 m) plots every 50 m, within which we centrally located a smaller (0.5 m x 0.5 m) quadrat ('small quadrat'). We first assessed the relative abundance (i.e., number of individuals) and diversity (i.e., species richness) of groundstorey (grasses and forbs) in the 5 m by 5 m quadrats, and the cover of woody plants (trees, shrubs, subshrubs) at 100 points, located every 2 m along the 200 m transect using a point-intercept method. We then conducted multiple animal surveys. Bird surveys, diurnal reptile searches and incidental records of vertebrates were conducted in an area 100 m x 200 m along transects. Bird surveys were conducted during spring to early summer (September to November) over two consecutive years. In each year, all sites were sampled twice for 20 minutes, on different days at different times, by a single observer. Surveys commenced from dawn and concluded by 12 noon, or earlier if ambient temperatures reached 30° C or if it became excessively windy (> 39 km/hr). Along the 200m transect, we positioned five grids of wet-pitfall traps at 50 m intervals for invertebrate sampling, and four trap lines for vertebrate sampling at the 50 m, 100 m, 150 m, and 200 m transect locations. Small mammals and reptiles were surveyed using dry pitfall traps, funnel traps, Elliott traps and timed searches. Vertebrate trap lines consisted of two 20 L buckets (150 mm deep), two 150 mm diameter PVC pipes (500-600 mm deep), and four double-ended funnel traps placed under or along a 20 m drift-fence. Dry pitfall traps were placed flush with the ground under the drift fence. Captured specimens were provided with shade cloth sheets, PVC tubes, Styrofoam blocks, litter and some soil in each trap to prevent over-heating or drowning in the event of rain. Ant rid powder and sprays were used at sites where ants were abundant. Funnel traps were located at either side of the drift fence, between the end pairs of pitfall traps. A 90% shade-cloth cover was placed over the top of the funnel traps to buffer temperatures inside the traps. Captured specimens were

provided with a cardboard roll and/or a sheet of sarking for shelter. In addition, four Elliot traps were also positioned near each trap line in appropriate habitat patches such as under shrubs, or near logs or rocks to enhance capture rates. Each trap was baited with a mixture of rolled oats and peanut butter. Traps were covered with shade cloth cover to buffer temperature extremes for captured specimens. Wet pitfall traps were 250 ml plastic screw-top containers half filled with ethylene glycol, installed at each corner of a 5 m x 5 m plot, plus one trap located centrally within the plot. Each pitfall trap was placed flush with the ground with a cover to prevent damage or loss of material due to rainfall. Traps were left open for five consecutive nights at each site. All vertebrate trap-lines were checked and cleared early each morning and late each afternoon over a 4 day period (eight times) and each species identified. Finally, two 30-minute habitat searches were undertaken at each 100 m x 200m site on different afternoons. Searches were targeted towards potential reptile habitat (e.g. open patches, leaf litter, logs, rocks, bark) by experienced personnel. Ground-dwelling invertebrates were sampled using both wet and dry pitfall traps with incidental specimens of large invertebrates (i.e. scorpions, spiders, centipedes, beetles, etc. > 1 cm, but excluding ants) collected from the vertebrate fauna pitfall traps each morning. All fauna surveys were conducted with approval from the New South Wales Animal Ethics Committee (Approval number: 140602/02).

Appendix S2. Molecular analyses conducted to characterize the soil microbial and animal community used to build our soil microbe-animal network (continental scale).

All soil DNA was extracted in triplicate, according to the methods employed by the Earth Microbiome Project (Bissett et al. 2016). Amplicons targeting the bacterial 16S rRNA, fungal Internal transcribed spacer (ITS) and Eukaryotic 18S rRNA genes were sequenced using the Illumina Miseq platform and the 27F – 519R, ITS1F–ITS4 and Euk_1391f–EukBr primer set, respectively (Bissett et al. 2016). Bioinformatic analyses were performed using MOTHUR (v1.34.1) as explained in Bissett et al. (2016). Operational Taxonomic Units (OTU) were picked at 97% sequence similarity. The OTU abundance tables were rarefied at 14237, 2901 and 4866 sequences/sample for bacteria, fungi and eukaryotes to ensure even sampling depth. In the case, of eukaryotes, we removed all fungal OTUs from the eukaryotic dataset as we are already using a higher resolution marker (ITS) to characterize the fungal community in our samples.

Appendix S3. Soil properties and current management

Soil properties were used using standardized lab protocols. Soil properties were measured as described in Eldridge et al. (2016) and Bissett et al. (2016). For the plant-animal dataset, we did not have soil pH. Soil pH information was obtained from Hengl et al. (2017) for the locations in this study. These authors produced 250m resolution global maps that included information on multiple soil properties. Predicted information on soil pH was cross-validated using the continental Australia dataset described above. In this dataset, pH measured in the field was significantly and positively related (Spearman $\rho = 0.65$; $P < 0.01$) to pH obtained for each plot using map predictions from Hengl et al. (2017).

For current management in the regional Australian dataset of plant-animal networks, information on the intensity of grazing by cattle was measured in the field. In brief, within the large quadrats used for the vegetation survey, we counted dung events, i.e. we considered a number of small fragments to have originated from one dung event, if the fragments were within an area of a few metres. We used algorithms, developed previously for the study area (Eldridge et al. 2016), to calculate the total oven-dried mass of dung per hectare based on the number of pellets recorded in the field. This total oven dried mass of dung was used as our measure of recent grazing intensity by cattle. Dung and pellet counts have been used widely to estimate the abundance of large herbivores (Johnson and Jarman 1987; Marques et al. 2001).

For the continental Australian dataset (BASE project) used to estimate soil microbe-animal networks, we obtained information on the density of cattle from Robinson et al. (2014). These authors produced 1 km resolution global maps that included information on livestock densities. Predicted information on cattle density from these models was cross-validated using the regional NSW dataset described above. In this dataset, cattle density measured in the field using the dung approach was significantly and positively related (Spearman $\rho = 0.20$; $P < 0.01$) to density of cattle obtained for each plot using map predictions from Robinson et al. (2014).

Table S1. List of taxa included in each of the ecological clusters within our aboveground and belowground networks and associated functional fungal traits.

Table S1 is available online as a Separate .XLS file under the Supporting Materials for this article.

Table S2. Correlation (Spearman) between relative abundance of ecological clusters used in this manuscript with the same clusters calculated after using centered log-ratio transformation.

| | Parameter | #0 | #1 | #2 | #3 | #4 | #5 | #6 |
|--------------------|-----------|--------|--------|--------|--------|--------|--------|--------|
| Aboveground | ρ | 0.95 | 0.96 | 0.90 | 0.99 | 0.97 | 0.96 | |
| | P-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | |
| Belowground | ρ | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.97 | 0.98 |
| | P-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |

Table S3. P-values associated to the relative contribution of the different predictors used to model the relative abundance of ecological clusters within our aboveground and belowground networks. AG = Aboveground network. BG = Belowground network.

| Dataset | Ecological cluster | Climatic legacies | Current climate | Management | Soil properties |
|----------------------------|--------------------|-------------------|-----------------|------------|-----------------|
| Aboveground network | AG#0 | 0.001 | 0.001 | 0.231 | 0.001 |
| | AG#1 | 0.001 | 0.001 | 0.879 | 0.001 |
| | AG#2 | 0.002 | 0.001 | 0.448 | 0.001 |
| | AG#3 | 0.289 | 0.104 | 0.841 | 0.026 |
| | AG#4 | 0.001 | 0.001 | 0.819 | 0.001 |
| | AG#5 | 0.001 | 0.001 | 0.085 | 0.001 |
| Belowground network | BG#0 | 0.001 | 0.001 | 0.459 | 0.001 |

| | | | | | |
|--|-------------|-------|-------|-------|-------|
| | BG#0 | 0.001 | 0.001 | 0.005 | 0.001 |
| | BG#1 | 0.001 | 0.001 | 0.001 | 0.001 |
| | BG#2 | 0.001 | 0.001 | 0.192 | 0.001 |
| | BG#3 | 0.001 | 0.001 | 0.121 | 0.001 |
| | BG#4 | 0.001 | 0.001 | 0.105 | 0.005 |
| | BG#5 | 0.001 | 0.001 | 0.001 | 0.001 |
| | BG#6 | 0.001 | 0.001 | 0.001 | 0.001 |

Table S4. Correlation (Spearman) between Random Forest importance across predictors calculated for each ecological cluster using the rfPermute and gradientforest R packages.

| Parameter | | #0 | #1 | #2 | #3 | #4 | #5 | #6 |
|--------------------|---------|--------|--------|--------|--------|--------|--------|--------|
| Aboveground | ρ | 0.87 | 0.89 | 0.91 | 0.85 | 0.94 | 0.92 | |
| | P-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | |
| Belowground | ρ | 0.90 | 0.93 | 0.85 | 0.91 | 0.91 | 0.95 | 0.91 |
| | P-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |

827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851

Table S5. Standardized direct effects ($0.01 < P < 0.05$) from the SEM in Fig. 2.

| Network | Response variables | | Predictors | Standardized effect | P-value |
|---------|--------------------|--|------------|---------------------|---------|
|---------|--------------------|--|------------|---------------------|---------|

**Aboveground
network**

| | | | | |
|--------------|---|--------------|--------|-------|
| AG#2 | ← | Latitude | 1.523 | 0.01 |
| AG#4 | ← | MAXT | -1.573 | 0.011 |
| AG#2 | ← | TSEA legacy | -0.701 | 0.015 |
| MINT legacy | ← | Latitude | -0.238 | 0.021 |
| AG#5 | ← | MAP | 0.863 | 0.023 |
| AG#1 | ← | MDR | -0.742 | 0.025 |
| AG#4 | ← | MAP | -0.616 | 0.027 |
| AG#2 | ← | MAXT legacy | -0.91 | 0.027 |
| AG#0 | ← | Latitude | -1.371 | 0.028 |
| AG#0 | ← | MINT | 1.493 | 0.03 |
| AG#1 | ← | TSEA legacy | 0.651 | 0.038 |
| AG#4 | ← | Sand content | 0.209 | 0.038 |
| AG#5 | ← | Sand content | 0.18 | 0.038 |
| AG#0 | ← | MAXT legacy | -2.011 | 0.042 |
| AG#1 | ← | MAXT legacy | 1.45 | 0.044 |
| Sand content | ← | PSEA | 0.31 | 0.045 |
| AG#0 | ← | Sand content | -0.158 | 0.046 |
| AG#1 | ← | Sand content | 0.205 | 0.047 |
| AG#5 | ← | MINT | -1.07 | 0.049 |

**Belowground
network**

| | | | | |
|--------|------|--------------|--------|-------|
| pH | <--- | MINT | 1.718 | 0.01 |
| BG#5 | <--- | Sand content | 0.101 | 0.01 |
| pH | <--- | TSEA | 1.046 | 0.011 |
| BG#6 | <--- | PSEA | -0.246 | 0.011 |
| BG#3 | <--- | MAP | 0.219 | 0.012 |
| BG#0 | <--- | TSEA legacy | 0.851 | 0.012 |
| Soil C | <--- | PSEA legacy | -0.232 | 0.013 |
| Soil P | <--- | MINT legacy | -0.597 | 0.015 |
| Soil P | <--- | Cattle | 0.101 | 0.015 |
| BG#0 | <--- | TSEA | -1.117 | 0.015 |
| BG#5 | <--- | TSEA | -0.88 | 0.016 |
| BG#5 | <--- | MINT legacy | -0.57 | 0.016 |
| pH | <--- | Latitude | 0.457 | 0.019 |
| BG#5 | <--- | MAP legacy | 0.194 | 0.019 |
| BG#5 | <--- | Longitude | 0.203 | 0.028 |

| | | | | |
|--------------|------|--------------|--------|-------|
| BG#4 | <--- | PSEA legacy | -0.335 | 0.028 |
| BG#1 | <--- | Longitude | -0.232 | 0.031 |
| BG#1 | <--- | Sand content | 0.108 | 0.033 |
| BG#3 | <--- | Soil P | -0.079 | 0.033 |
| pH | <--- | TSEA legacy | -0.704 | 0.034 |
| BG#4 | <--- | MDR | -1.274 | 0.034 |
| BG#5 | <--- | MDR | -0.595 | 0.034 |
| BG#4 | <--- | MAXT | 2.658 | 0.035 |
| MAP legacy | <--- | Longitude | 0.165 | 0.036 |
| MAP legacy | <--- | Latitude | 0.142 | 0.037 |
| TSEA | <--- | Latitude | -0.144 | 0.038 |
| pH | <--- | Cattle | 0.104 | 0.038 |
| Sand content | <--- | PSEA legacy | -0.213 | 0.039 |
| BG#2 | <--- | MAP | 0.134 | 0.043 |
| BG#6 | <--- | Cattle | 0.148 | 0.044 |
| PSEA legacy | <--- | Latitude | -0.096 | 0.049 |

852

853

854

855

856

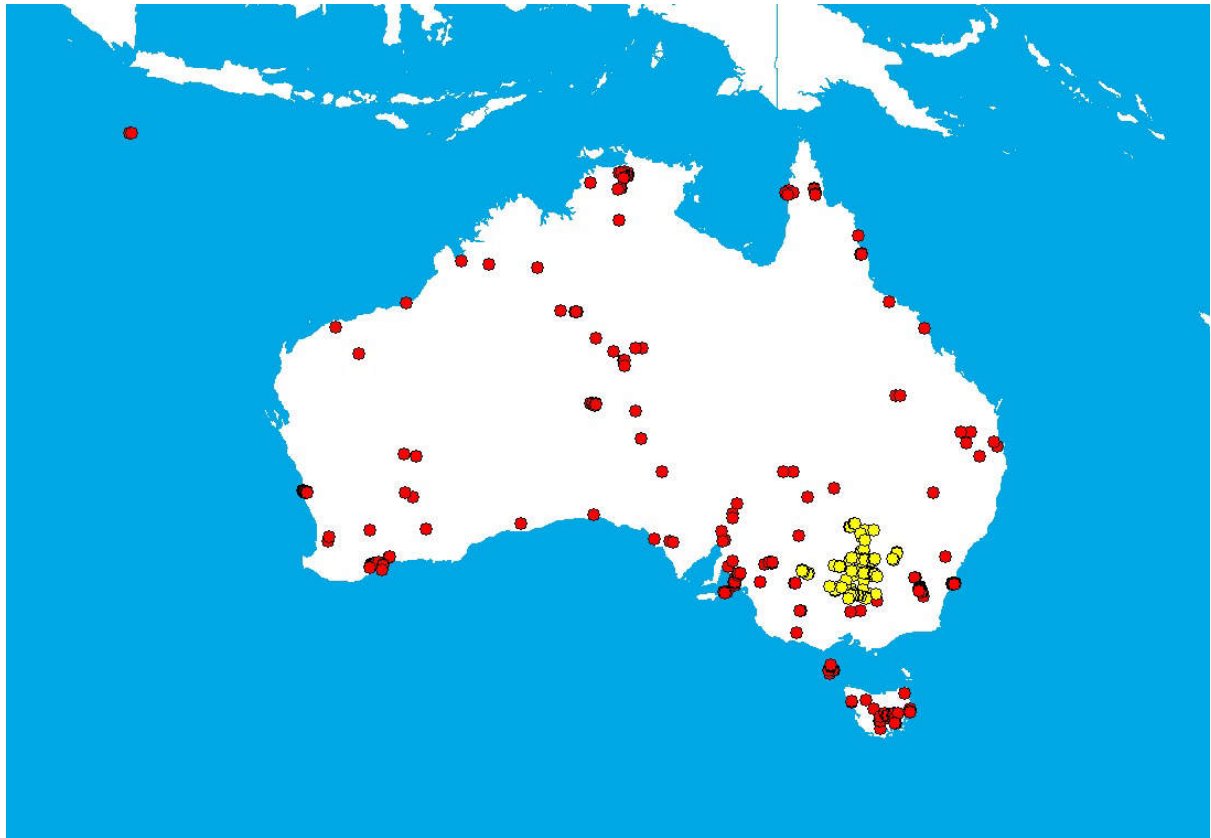
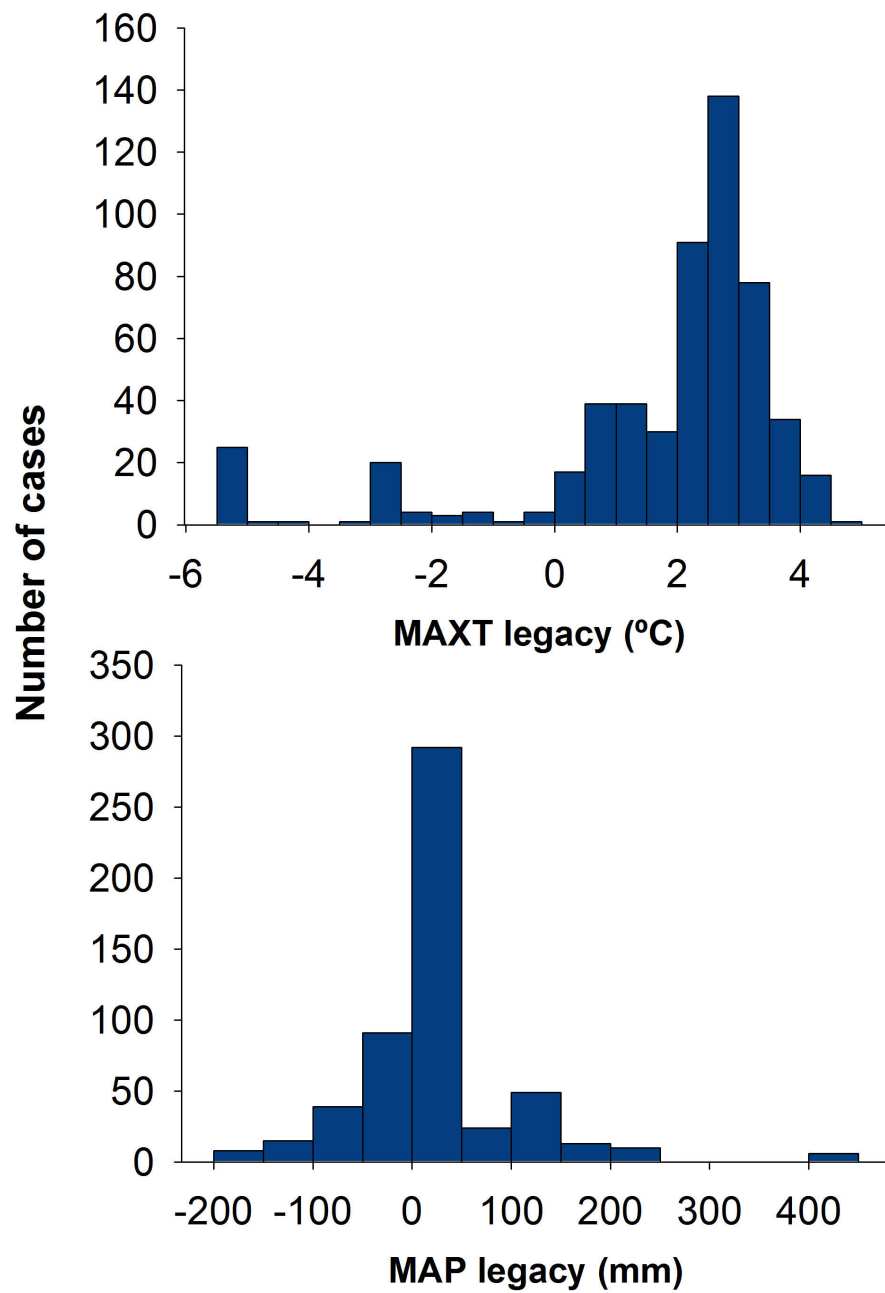


Figure S1. Location of the sites included in the studies of aboveground network ($n = 108$) in yellow and the belowground network ($n = 439$) in red.

872



873

874

875 **Figure S2.** MAP and MAXT legacy distribution across the 547 locations included in this
876 study.

877

878

879

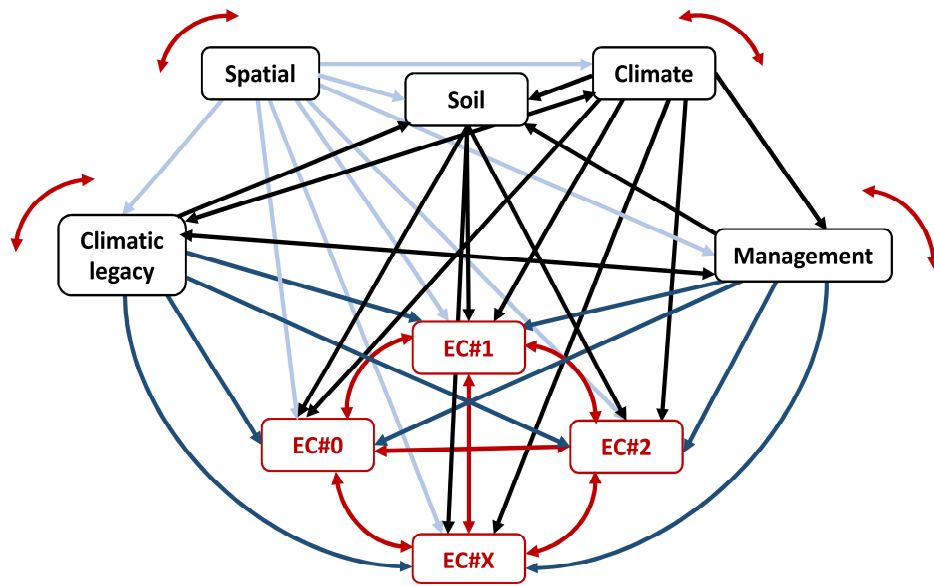
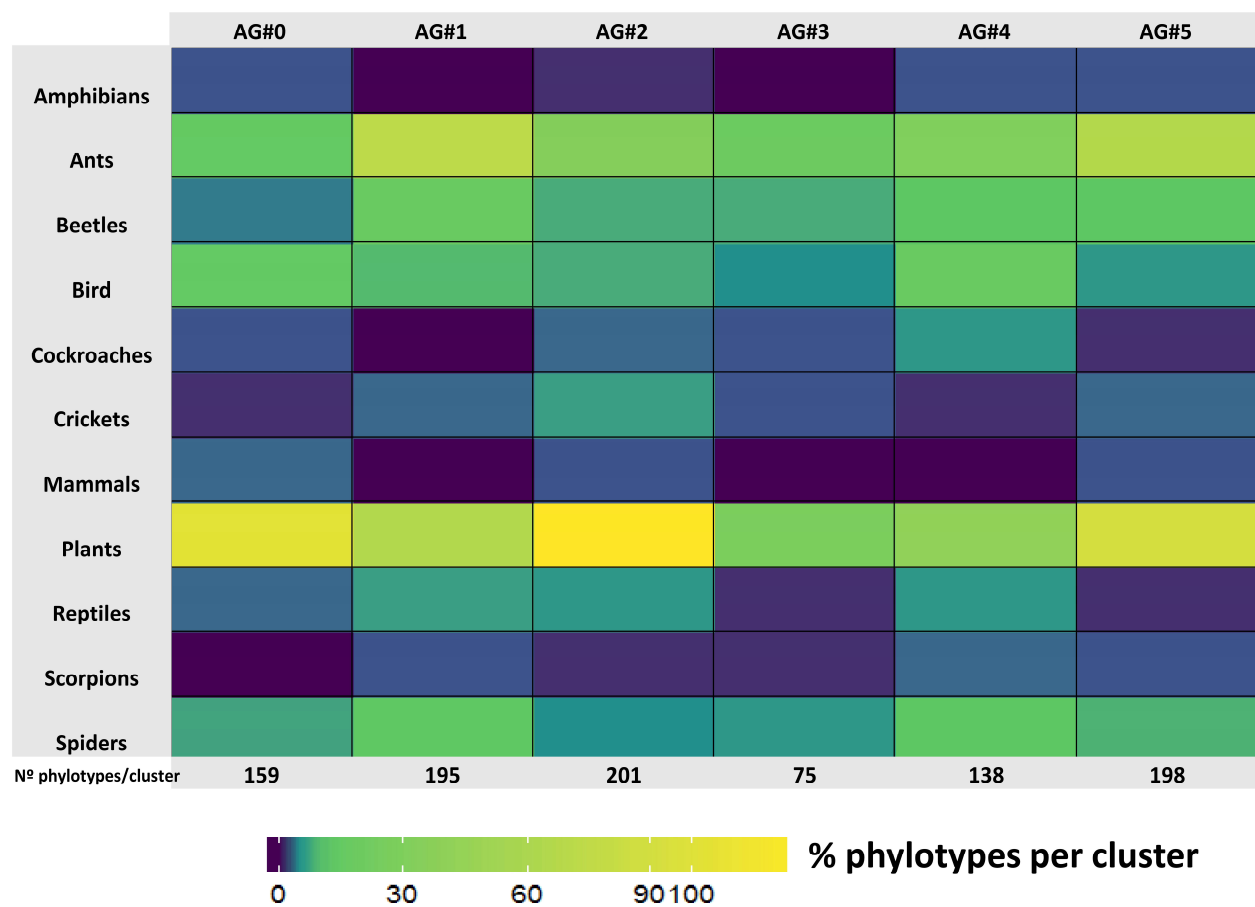


Figure S3. *A priori* structural equation model including direct and indirect effects of geographical location, climatic legacies, current climate and management on the relative abundance of ecological clusters (EC #) within our aboveground and belowground networks. Predictors within climatic legacy, current climate, spatial (latitude and longitude) and management are allowed to co-vary in these analyses.

899
900



901
902
903
904
905
906
907

Figure S4. Taxonomic composition (% of taxa within each ecological cluster) for six well-defined clusters of strongly co-occurring soil taxa within our aboveground. AG = Aboveground network.

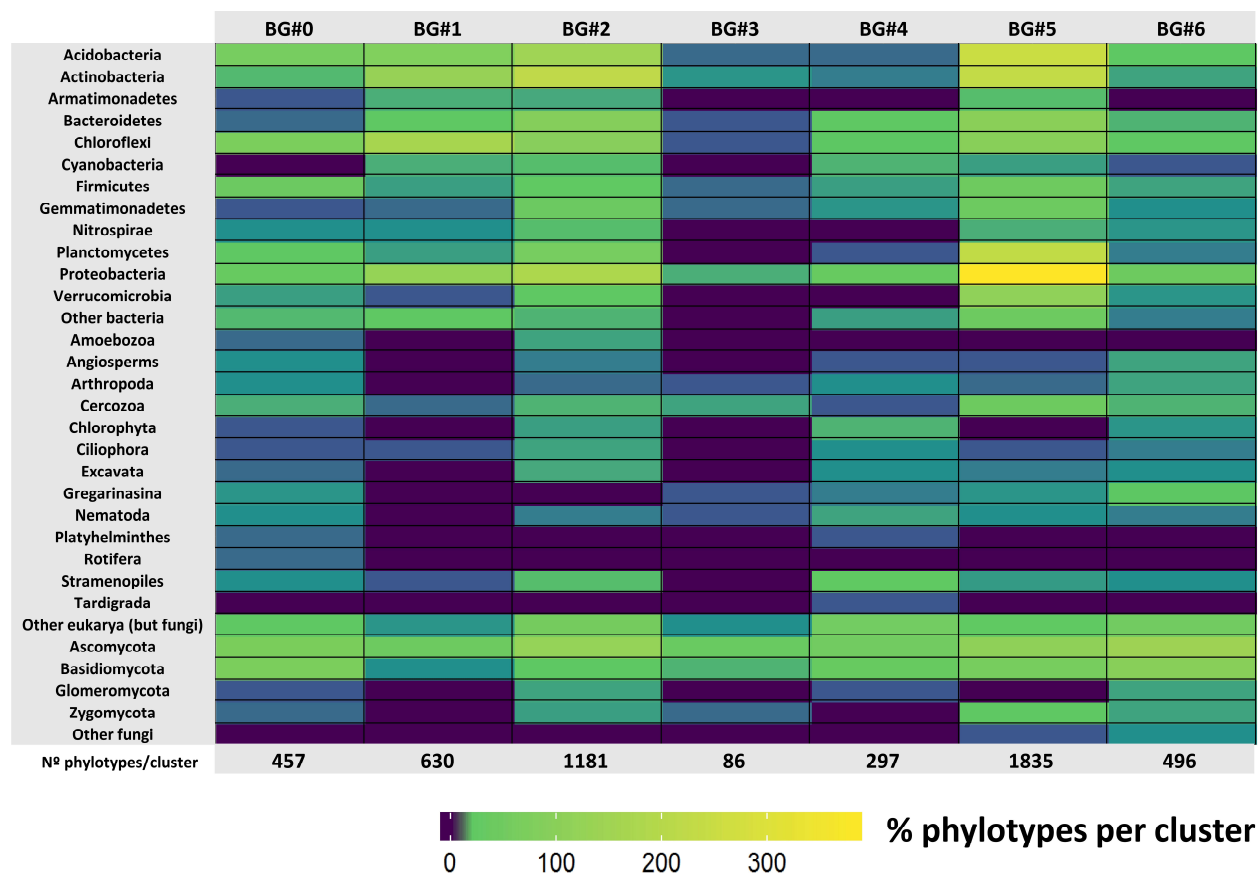
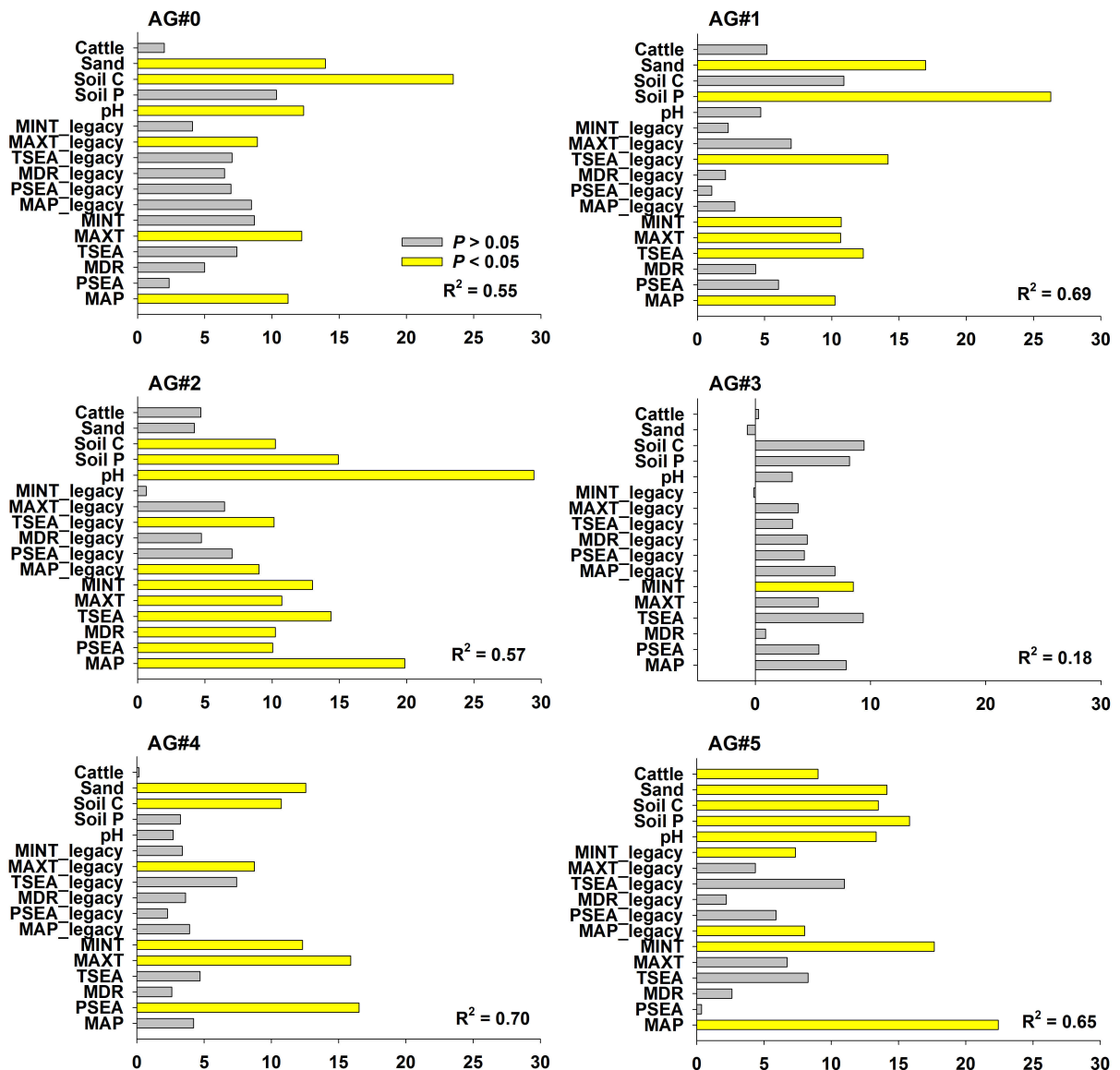


Figure S5. Taxonomic composition (% of taxa within each ecological cluster) for seven well-defined clusters of strongly co-occurring soil taxa within our belowground. BG = Belowground network.

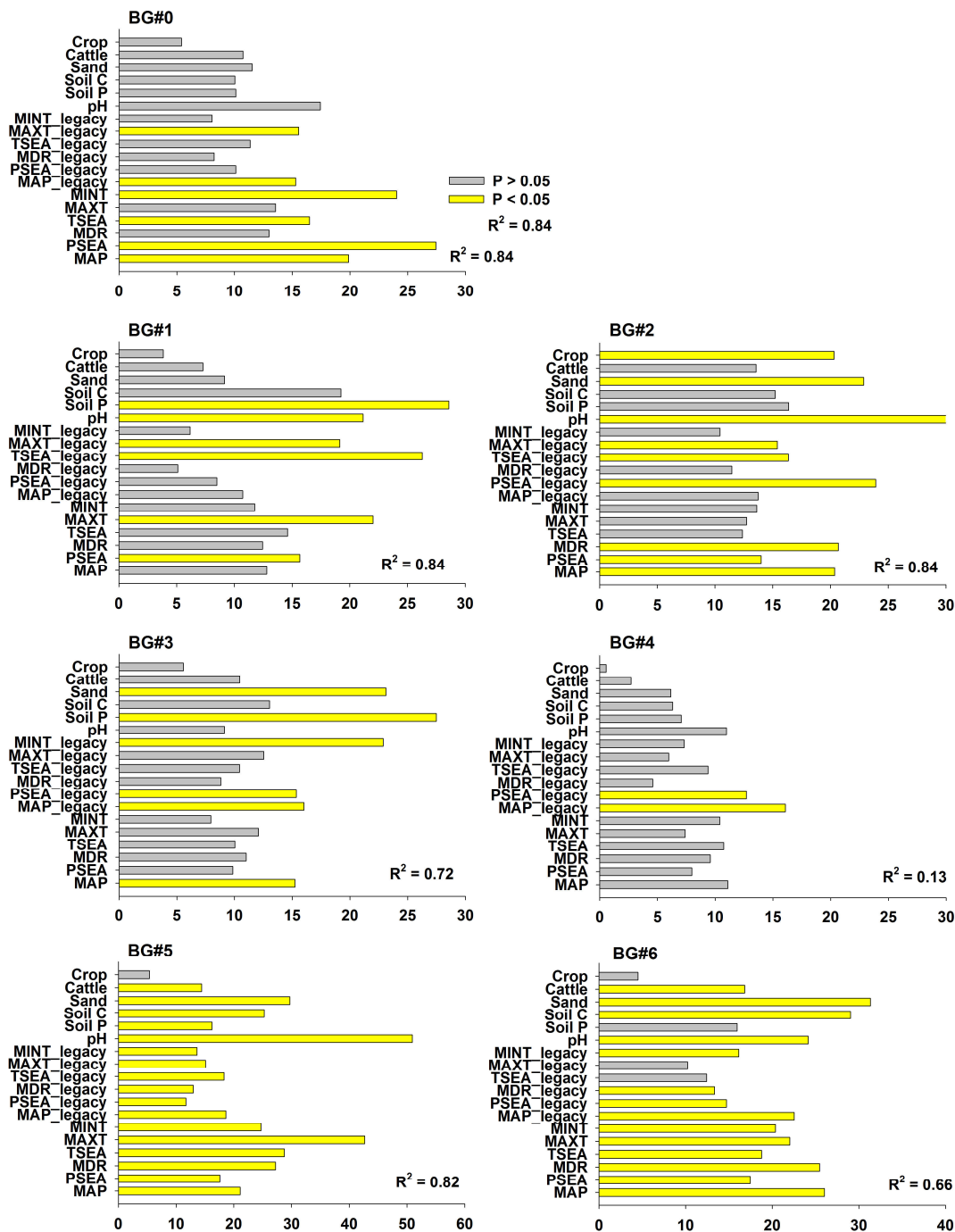
Aboveground network



Importance (% of increase in MSE)

Figure S6. Random Forest analysis (rfPermute R package) aiming to identify the best individual predictors of the relative abundance of ecological clusters within our aboveground network. Predictors include those within climatic legacies, current climate and management categories. MSE = Mean Square Error. AG = Aboveground network.

Belowground network



919

Importance (% of increase in MSE)

920

Figure S7. Random Forest analysis (rfPermute R package) aiming to identify the best individual predictors of the relative abundance of ecological clusters within our belowground network. Predictors include those within climatic legacies, current climate and management categories. MSE = Mean Square Error. BG = Belowground network.

923

Aboveground network

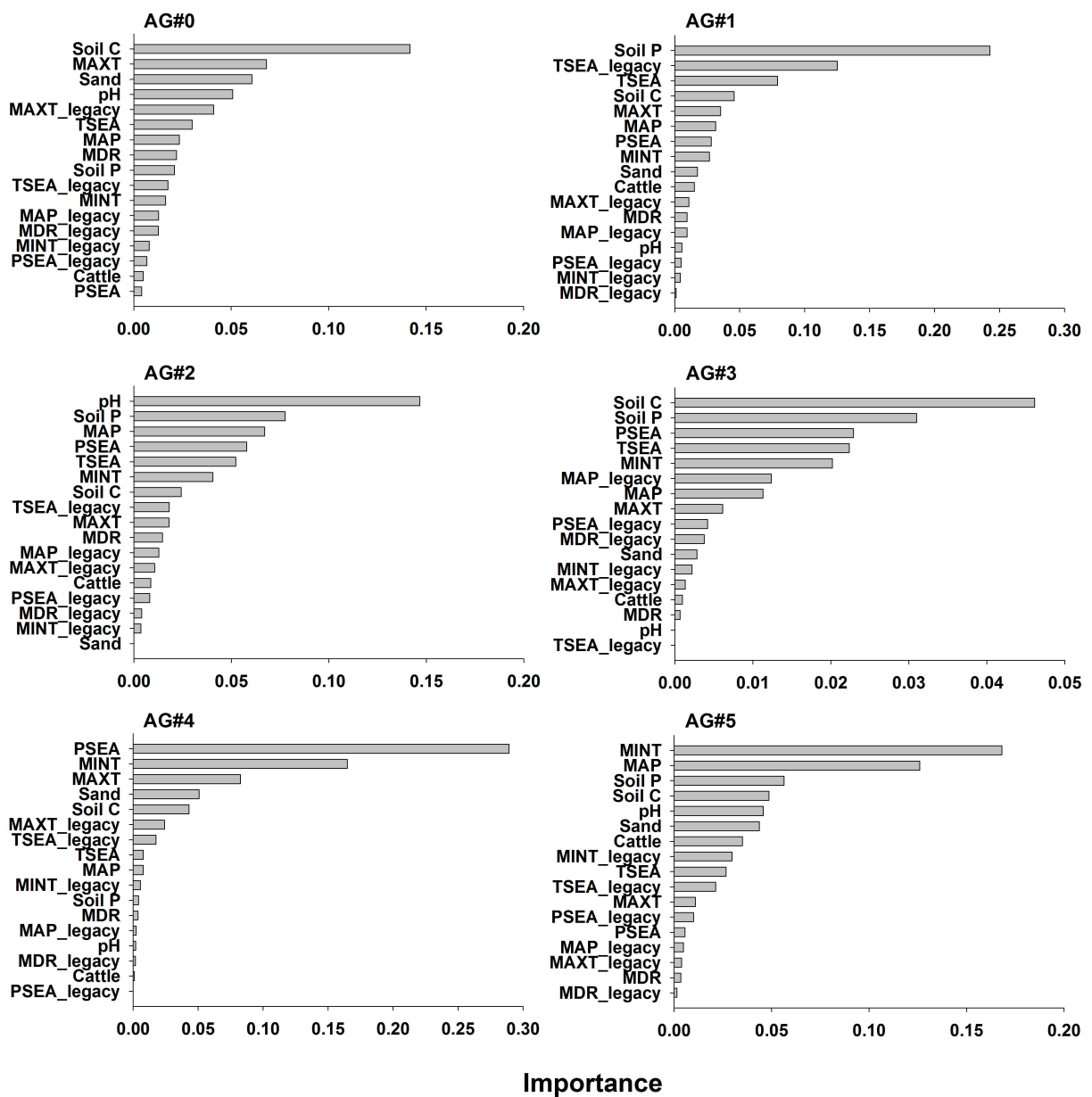
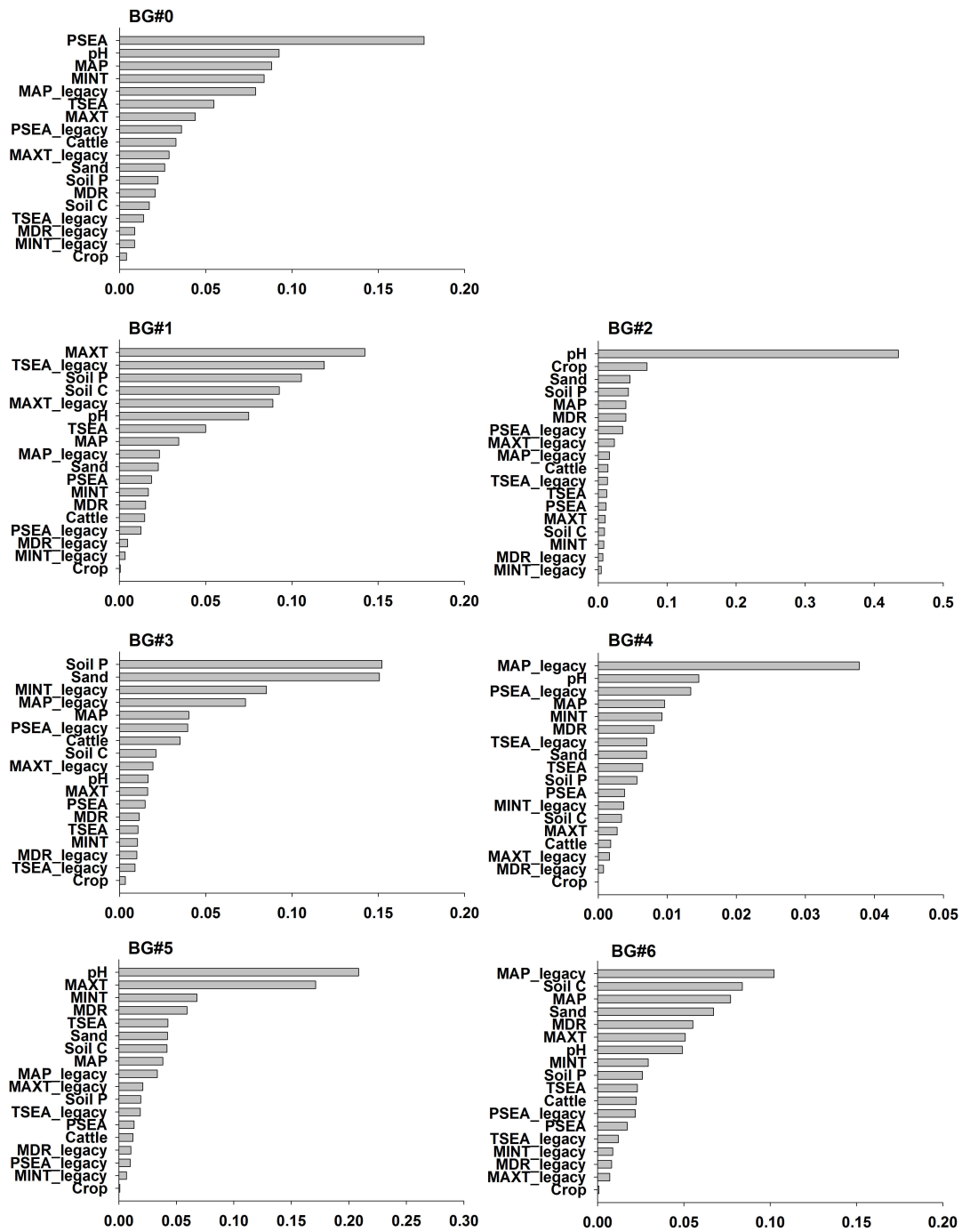


Figure S8. Random Forest analysis (gradientforest R package) aiming to identify the best individual predictors of the relative abundance of ecological clusters within our aboveground network. Predictors include those within climatic legacies, current climate and management categories. AG = Aboveground network.

Belowground network



Importance

Figure S9. Random Forest analysis (gradientforest R package) aiming to identify the best individual predictors of the relative abundance of ecological clusters within our belowground network. Predictors include those within climatic legacies, current climate and management categories. BG = Belowground network.

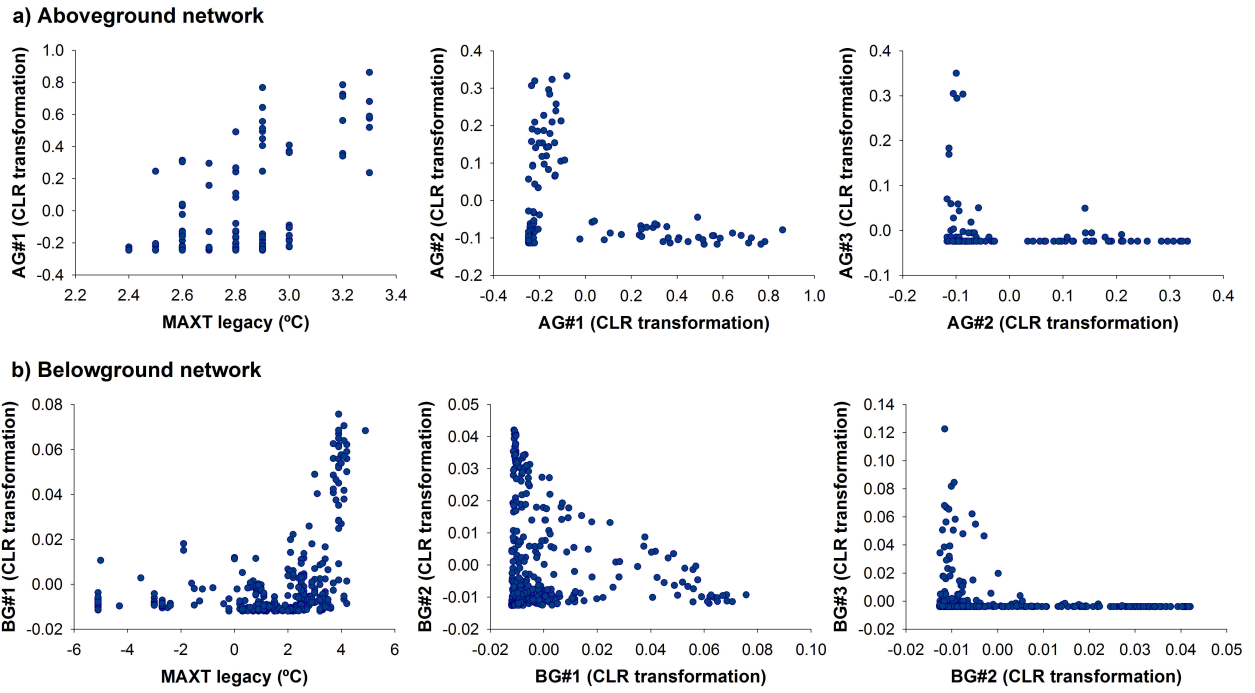


Figure S10. Selected relationships from our SEMs. Panels include relationship between maximum temperature legacy and ecological clusters #1 within our aboveground and belowground networks. Also, selected relationships between ecological clusters #1 and 2, and ecological clusters #2 and 3 in both independent networks. Data was centered log-ratio transformed before ecological clusters were calculated. AG = Aboveground network. BG = Belowground network. CLR = Centered log-transformation.

References (not listed in the main text)

- Eldridge, D.J., Delgado-Baquerizo, M., Travers, S.J., Val, J. & Oliver, I. (2016). Do grazing intensity and herbivore type affect soil health? Insights from a semi-arid productivity gradient. *J. Appl. Ecol.* **54**, 976-985.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **12**, e0169748.
- Johnson, C.N. & Jarman, P.J. (1987). Macropod studies at Wallaby Creek. VI: A validation of the use of dung-pellet counts for measuring absolute densities of populations of macropodids. *Aust. Wildl. Res.* **14**, 139-146.
- Marques, F.F.C., Buckland, S.T., Goffin, D., Dixon, C.E., Borchers, D.L., Mayle, B.A. et al. (2001). Estimating deer abundance from line transect surveys of dung: sika deer in southern Scotland. *J. Appl. Ecol.* **38**, 349-363.
- Robinson, T.P., Wint, G.R.W., Conchedda, G., Van Boeckel, T.P., Ercoli, V., Palamara, E., et al. (2014). Mapping the global distribution of livestock. *PLoS ONE* **9**, e96084.