Ecography

Supplementary material

Appendix 1 Additional methodological information

Study site

The Azorean archipelago is located in the North Atlantic, roughly between 37º to 40º N latitude and 25º to 3º W longitude. Since human occupation of the islands, there have been substantial changes in the size and quality of native habitats, mostly due to the creation of urban areas, agriculture fields, pastures and non-native plantations (Borges et al. 2008, Triantis et al. 2010). Terceira Island (Fig. 1), a roughly circular island of 402 km^2 area, contains the largest area of native pristine forests in the Azores (Triantis et al. 2010), with five main fragments of native forest distributed across four main volcanic polygenetic complexes.

Data collection

Arthropods were sampled using nine 50m x 50m plots located in four native fragments of pristine forest on Terceira setup within the ISLANBIODIV project (Cicconardi et al. 2017): Serra de Santa Bárbara (Plots T07, T48, T49, T164), Biscoito da Ferraria (Plots T01, T02, T41), Terra Brava (Plot T15) and Galhardo (Plot T33) (see Figure S1, below). Arthropods were sampled using a passive flight interception trap called a SLAM (Sea, Land, and Air Malaise) trap. The SLAM traps are approximately 110 x 110 x 110 cm (see Fig. S2 in Appendix S1), and work by intercepting arthropods on an area of black mesh and funnelling them into a sampling bottle filled with a killing liquid. Propylene-glycol was used as it persists for a long time without evaporating, and enables the collection of good quality specimens for posterior DNA extraction. The collecting bottles were collected and changed every three months; thus, each sample covers one season of the year. However, due to logistical challenges and problems with the traps, in a very small number of cases (see Table S1 in Appendix S1) certain three-month samples were excluded. Further details on the method can be found in Borges et al. (2017). Samples were sorted and individuals identified to species level by experienced taxonomists and to morpho species where specific determinations were not possible. The arthropods were then grouped by their native and non-native colonisation strategies. For the current study, we used data sampled over the years 2012 –2016 (inclusive). Sampling only commenced half way through 2012 (see Table S1, below) and thus for the temporal beta-diversity analyses we used 2013 as the base year to ensure that the pairwise comparisons were based on equal samples. However, for the species-

level turnover analyses we used the full dataset as the missing months here are less likely to affect the results.

For the climatic data, we used the CIELO Model (Azevedo et al. 1999). The CIELO model is a simple parcel model, based on the transformations experienced by an air mass ascending a mountain, which simulates the evolution of an air parcel's physical properties, starting from the sea level. CIELO climatic data for the sites were extracted from Borges et al. (2006). In addition, we used an 'index of Disturbance' developed by Cardoso et al. (2013) for the Azores that models disturbance by considering landscape configuration and proximity of human-modified habitat to each patch of pristine native forest. Thus, in this study we represent anthropogenic disturbance of a plot by explicitly considering landscape configuration and the amount of neighboring anthropogenic habitats, ranked according to their level of intensity-use.

Species sorting and classification

Parataxonomists sorted samples to orders, and posteriorly to Recognizable Taxonomic Units (RTUs). One of the authors (PAVB) then identified to species level the RTUs of the following arthropod orders: Diplopoda (Chordeumatida, Julida), Chilopoda (Geophilomorpha, Lithobiomorpha, Scolopendromorpha), Arachnida (Araneae, Opiliones, Pseudoscorpiones) and Insecta (Blattaria, Coleoptera, Hemiptera, Microcoryphia, Neuroptera, Psocoptera, Thysanoptera, Trichoptera). All material is stored at EDTP—Entomoteca Dalberto Teixeira Pombo, University of Azores, Angra do Heroísmo, Portugal.

Arthropods were grouped into three colonization categories: endemic (i.e. restricted to the Azores); native non-endemic, i.e. species that arrived naturally to the archipelago but are present both in the Azorean Islands and elsewhere; and non-native species, i.e., species whose original distribution range did not include the Azores and that are believed to have been introduced in the Macaronesian region after human settlement in the 15th century. The non-native status was inferred either from historical records of detected species introductions or from their current distribution being closely associated with human activity. For unidentified species, if other species in the same genus, subfamily or family were present in the archipelago and all belonged to the same colonization category (according to Borges et al. 2010), the unknown species were classified similarly. Otherwise, we assumed the species to be native. For simplicity, endemic and native non-endemic species were grouped and are termed "native" throughout the text.

Sensitivity analyses methodology

To ensure our sampling was sufficient we calculated sampling completeness estimates for each year in each plot using the iNEXT R package (Hill number order q = 0, Hsieh et al. 2016).

Due to the resources required to sample arthropods in multiple plots every three months over five years, we only placed one SLAM trap in each plot. To determine whether this was sufficient to capture the relevant properties (e.g. species composition, distribution of abundance) of the sampled communities, in 2015 we set up a tenth plot in Terra Brava (Plot T18), in which we placed three SLAM traps, spaced 30 m apart, and sampled every month. We constructed three overall assemblage matrices whereby, for each SLAM trap, the data from the 12 months of 2015 were pooled into one sample. To determine whether community composition was similar across the three traps, we calculated a number of summary statistics: 1) the number of species in each assemblage, 2) the total abundance of each assemblage, 3) the proportion of species that were sampled in all three traps (i.e. are present in all three assemblages), and 4) the distribution of abundance across the species in each assemblage. In regards to (4), we fitted the gambin species abundance distribution model, using the 'gambin' R package and a subsampling procedure to ensure sample sizes were kept consistent (see Matthews et al. 2014), to the abundances of the species in each assemblage and recorded the alpha (shape) parameter (Matthews et al. 2014). Finally, to check that spatial species turnover across the three traps was low, we calculated multisite spatial beta-diversity across the three assemblages using presence-absence data (*'beta.multi'* function; Sorensen index) and abundance data (*'beta.multi.abund'* function; Bray-Curtis multiple-site dissimilarity), using the 'betapart' R package (Baselga et al. 2017). As most species were found in all three traps, it was not possible to use a null model approach that constrained the marginal totals. All analyses were undertaken using R (version 3.4.3, R Core Team 2017).

References

Azevedo, E. B. d. et al. 1999. Modelling the local climate in island environments: water balance applications. — Agricultural Water Management 40: 393-403.

Baselga, A. et al. 2017. betapart: partitioning beta diversity into turnover and nestedness components. R Package.

Borges, P. A. V. et al. 2006. Invasibility and species richness of island endemic arthropods: a general model of endemic vs. exotic species. — Journal of Biogeography 33: 169-187.

Borges, P. A. V. et al. 2017. Seasonal dynamics of arthropods in the humid native forests of Terceira Island (Azores). — Arquipelago Life and Marine Sciences 34: 105-122.

Borges, P. A. V. et al. 2010. List of arthropods (Arthropoda). — In: Borges, P. A. V. et al. (eds), A list of the terrestrial and marine biota from the Azores. Princípia, pp. 179-246.

Cardoso, P. et al. 2013. Integrating landscape disturbance and indicator species in conservation studies. — PLOS ONE 8: e63294.

Cicconardi, F. et al. 2017. MtDNA metagenomics reveals large-scale invasion of belowground arthropod communities by introduced species. — Molecular Ecology 26: 3104-3115.

Matthews, T. J. et al. 2014. The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation. — Ecography 37: 1002-1011.

R Core Team 2017. R: A language and environment for statistical computing. R foundation for statistical computing.

Triantis, K. A. et al. 2010. Extinction debt on oceanic islands. — Ecography 33: 285-294.
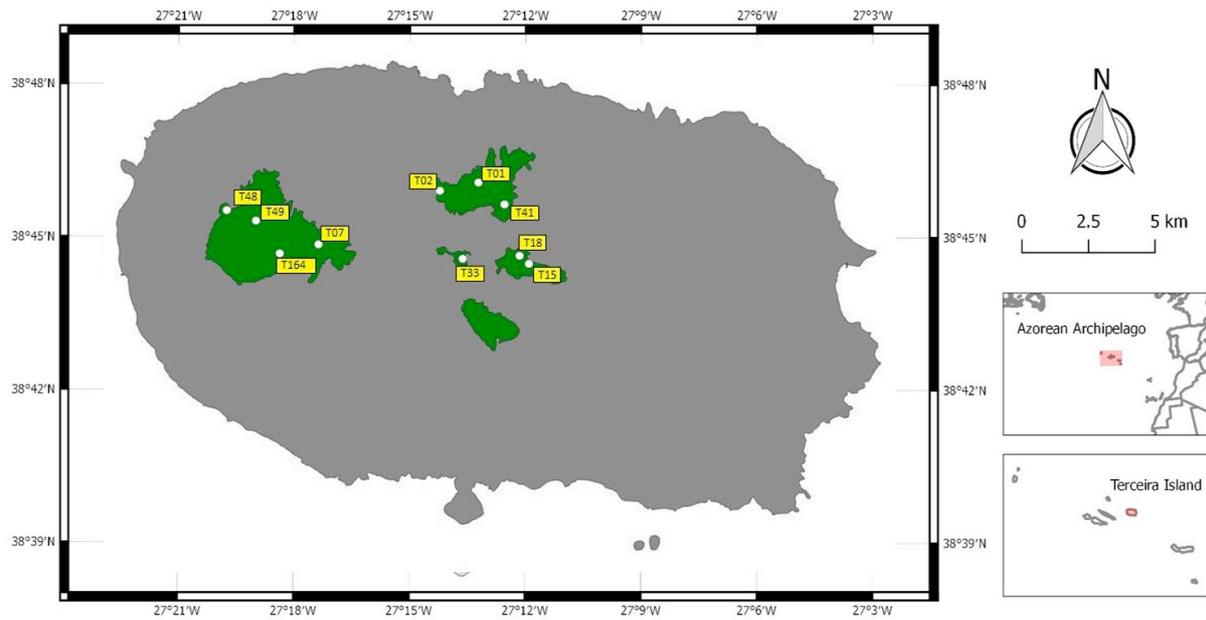
Plot locations

Figure A1. A map of the ten sampled 50m x 50m plots and areas of native forest on the island of Terceira, Azores.

Figure A2. Example SLAM trap in a plot within a native forest fragment on Terceira Island (Terra-Brava T18), Azores

Table A1. Sampling dates for the nine plots across the five years. A '1' indicates that sampling was undertaken in this plot during this time period, whilst a '0' indicates that sampling was not undertaken during this time period due to either logistical issues or problems with the traps. Codes as in Figure S1: 1- TER-NFBF-T-01; 2- TER-NFBF-T-02; 3 - TER-NFBF-TP41; 4- TER-NFPG-T-33; 5- TER-NFSB-T-07; 6- TER-NFSB-T164; 7-TER-NFSB-TE48; 8- TER-NFSB-TE49; 9- TER-NFTB-T-15,

Plot	2012 Sep.	2012 Dec.	2013 May	2013 Sep.	2013 Dec.	2014 March	2014 June	2014 Sep.	2014 Dec.
1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1
4	1	0	1	1	1	1	1	1	1
5	1	0	1	1	1	1	1	1	1
6	1	0	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1

Plot	2015 March	2015 Sep.	2015 June	2015 Dec.	2016 March	2016 June	2016 Sep.	2016 Dec.
1	1	1	1	1	1	1	1	1
2	1	1	1	0	1	1	1	1
3	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	0
6	1	0	1	1	1	1	1	1
7	1	1	1	0	1	1	1	1
8	1	1	1	1	1	0	1	1
9	1	1	1	1	1	1	1	1

Appendix 2 Temporal beta-diversity results using the Carvalho et al. partition framework
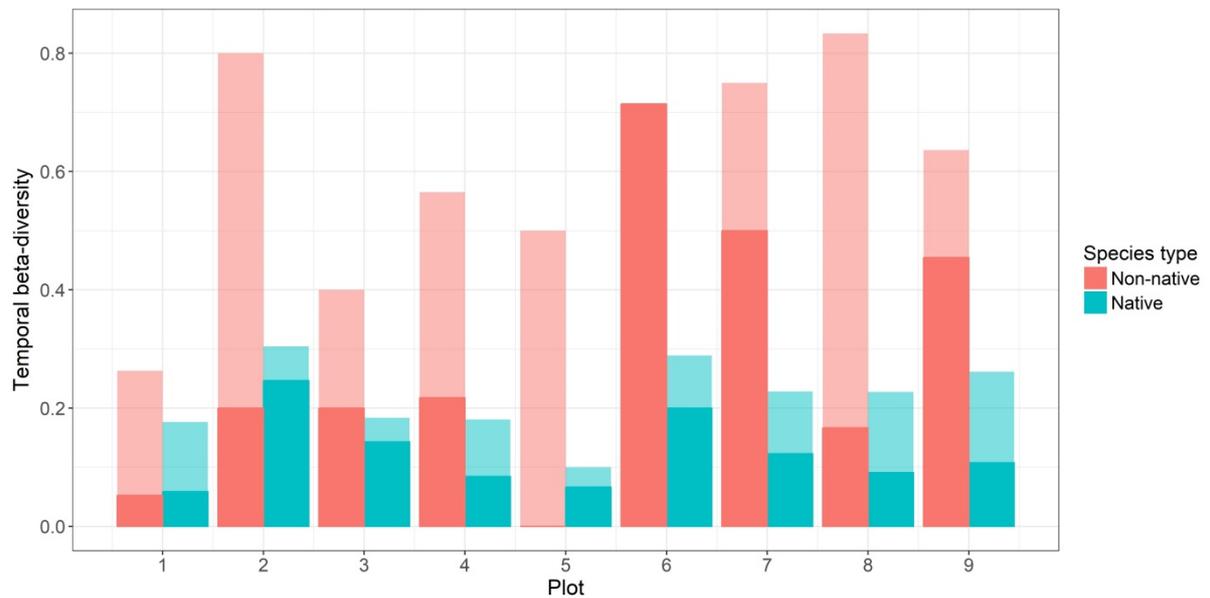

Figure A3. Pairwise temporal beta-diversity values for native and non-native arthropod species across nine plots of native forest on Terceira Island, in the Azores (Plot notation as in Table S1 and Figure S1). For each plot, temporal beta-diversity was calculated using the pooled 2013 samples and the pooled 2016 samples. For each plot, temporal beta-diversity was calculated separately for native species (blue bars) and non-native species (red bars). The height of each bar corresponds to the overall temporal beta-diversity. Overall temporal beta was also partitioned into richness differences and turnover components using the Carvalho partition approach (Carvalho et al. 2012). Within each bar, the dark shaded area rising from the x-axis corresponds to the richness differences component of overall temporal beta diversity. Thus, the lighter shaded area within each bar that rises from the dark shaded area corresponds to the turnover component.

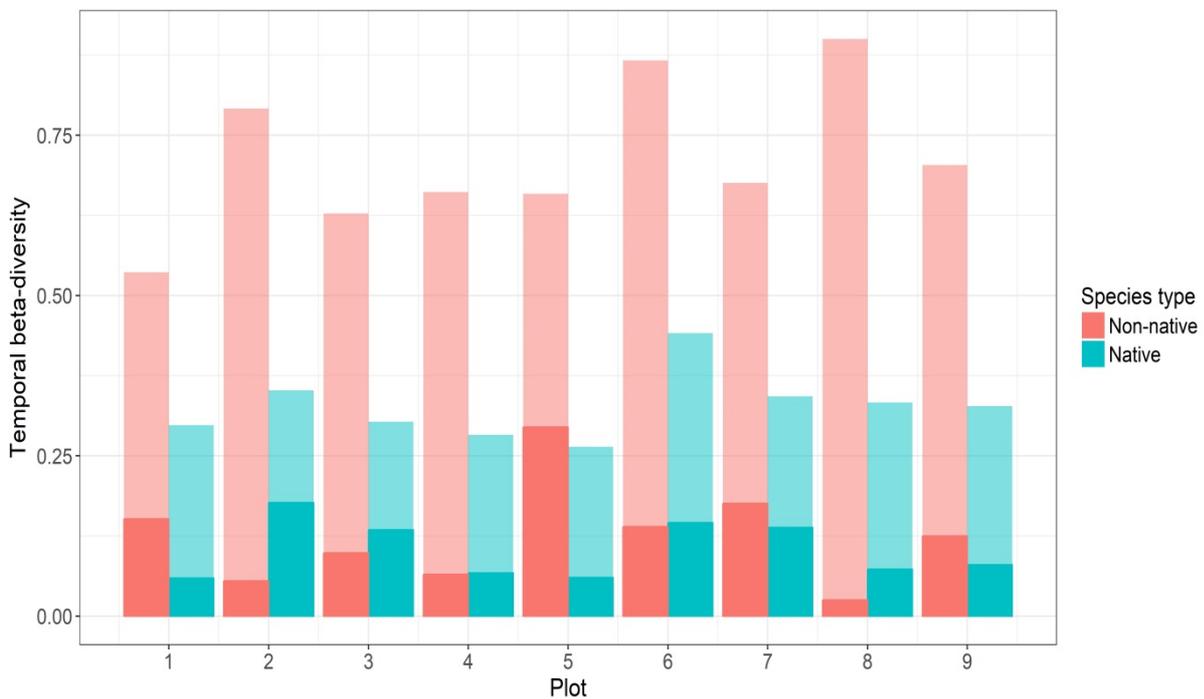**Appendix 3 Multiple-time dissimilarity results**

Figure A4. Multiple-time dissimilarity values for native and non-native arthropod species across nine plots of native forest on Terceira Island, in the Azores. For each plot, multiple-time dissimilarity was calculated across four years of data (2013 – 2016). For each plot, temporal beta-diversity was calculated separately for native species (blue bars) and non-native species (red bars). The height of each bar corresponds to the overall multiple-time dissimilarity was. Overall multiple-time dissimilarity was also partitioned into nestedness and turnover components using the Baselga partition approach (Baselga et al. 2017). Within each bar, the dark shaded area rising from the x-axis corresponds to the nestedness component of overall temporal beta diversity. Thus, the lighter shaded area within each bar that rises from the dark shaded area corresponds to the turnover component.

Appendix 4 Sensitivity Analyses Full Results

Table A2. Sampling coverage estimates for the nine plots across the five years. SC is the sampling coverage for the observed data, and the LCI and UCI are the 95% lower and upper confidence intervals, respectively.

SC	LCI	UCI	Plot	Year	SC	LCI	UCI	Plot	Year
0.992	0.987	0.996	1	2012	0.983	0.967	1.000	8	2012
0.986	0.981	0.991	1	2013	0.987	0.981	0.993	8	2013
0.992	0.987	0.997	1	2014	0.989	0.983	0.995	8	2014
0.987	0.982	0.993	1	2015	0.981	0.968	0.993	8	2015
0.986	0.981	0.992	1	2016	0.981	0.970	0.991	8	2016
0.942	0.911	0.974	2	2012	0.979	0.965	0.992	9	2012
0.984	0.977	0.992	2	2013	0.990	0.986	0.995	9	2013
0.985	0.975	0.995	2	2014	0.985	0.976	0.993	9	2014
0.967	0.952	0.981	2	2015	0.976	0.964	0.989	9	2015
0.989	0.983	0.995	2	2016	0.973	0.958	0.987	9	2016
0.970	0.956	0.985	3	2012					
0.990	0.982	0.997	3	2013					
0.996	0.993	1.000	3	2014					
0.972	0.959	0.985	3	2015					
0.985	0.974	0.996	3	2016					
0.918	0.878	0.957	4	2012					
0.991	0.987	0.995	4	2013					
0.991	0.988	0.994	4	2014					
0.979	0.972	0.986	4	2015					
0.985	0.978	0.992	4	2016					
0.810	0.666	0.954	5	2012					
0.994	0.989	0.998	5	2013					
0.989	0.982	0.996	5	2014					
0.984	0.976	0.992	5	2015					
0.988	0.978	0.999	5	2016					
0.902	0.845	0.958	6	2012					
0.954	0.931	0.977	6	2013					
0.968	0.956	0.980	6	2014					
0.965	0.947	0.984	6	2015					
0.966	0.949	0.984	6	2016					
0.983	0.957	1.009	7	2012					
0.977	0.965	0.990	7	2013					
0.985	0.977	0.993	7	2014					
0.963	0.943	0.984	7	2015					
0.910	0.867	0.953	7	2016					

Using three SLAM traps in a single plot

Using three SLAM traps in a plot rather than one did not result in substantially different sampled communities. Species richness across the three traps in T18 only ranged 49 – 51 species, whilst abundance varied from 1380 to 1739 individuals per trap. The alpha parameter of the gambin distribution was relatively constant between the three traps (2.47, 2.44 and 3.69), indicating similar SAD shapes. 58% of the total sampled species (n = 38) were present in all three traps and overall spatial beta-diversity was low based on both presence/absence data (Sorensen dissimilarity = 0.24) and abundance data (Bray-Curtis multiple-site dissimilarity = 0.31). Thus, we are confident that our general sampling protocol is sufficient to obtain representative samples of the arthropod fauna in each plot.