###Restrictions and embargos

1. In the Wakatobi dataset (O’Connell et al. 2020), we have removed one species from the published dataset to avoid publishing the islands on which it is found: *Cacatua sulphurea* – a critically endangered cockatoo that is at risk from the caged bird trade. Please contact Tom Martin (a paper co-author; tom\_martin\_2010@yahoo.co.uk) if this information is required.
2. The Zhoushan dataset (Zhao et al.) has only recently been developed through an extensive field work campaign. As such, this dataset has been embargoed for a maximum of one year to allow for it to be used in a different publication. The R scripts and data files will be updated when this embargo is lifted. However, note that results from running the analyses herein may differ slightly from those published given this omission. Please contact the lead author if this information is required.

####R code files

Within the R Code directory, there are four .R files:

1. DAR\_modelling – this is the main script for running the analyses, secondary analyses and making the figures. It uses parallel processing.
2. DAR\_SOURCE – this holds all of the functions that are used in DAR\_modelling.
3. DAR\_DATA\_SOURCE – loads in the consensus phylogeny, the trait data (for extant and extinct species) and the datasets. Depending on what settings are used in DAR\_modelling, uses body-size corrected traits or not, and the full datasets or the land bird versions.
4. HEATMAP – the code for making Figure 5.

#####Data files

A separate Dataset\_information word document is included as part of the repository which provides information for each dataset individually.

Within the Data directory, there are five .rds data files:

1. AllSP\_DATA – the full datasets (i.e. including all species) used in the main analyses. Stored as a list with 10 elements:

* 1st element = the filenames of the 50 datasets (51 after the embargo is lifted).
* 2nd = The presence-absence matrices (order matches that in the 1st element). The format of the datasets is a presence-absence matrix, with species as rows and sites as columns (this is transposed in the scripts). The bottom two rows include the island area and species richness (sp.r) totals for each island. All island areas are in hectares (which are converted to km2 in the scripts). All species names follow the Jetz et al (2012) BirdTree taxonomy. The first 25 datasets are habitat islands, with the remaining 25 (26 when the embargo is lifted) representing the true island datasets.
* 3rd = the filenames of the 10 datasets including alien species.
* 4th = the presence-absence matrices for the datasets including alien species (follows order of 3). Same format as 2, but here with alien species included.
* 5th = the filenames of the 8 datasets representing the historic period.
* 6th = the presence-absence matrices for the datasets representing the historic period (follows order of 5). Same format as 2, but here with historically extinct species included.
* 7th = the filenames of the 10 datasets from the prehistoric dataset analysis.
* 8th = the presence-absence matrices for the datasets from the prehistoric dataset analysis (follows order of 7). Same format as 2. Includes both the prehistoric period dataset, and the corresponding current assemblage dataset excluding marine and alien species. These can be distinguished by the filenames in 7 (containing either Modern or Prehistoric).
* 9th = combines 5 and 7 (filenames of all datasets used in the extinct species analyses).
* 10th = combines 6 and 8 (all datasets used in the extinct species analyses).

1. LANDBIRD\_DATA – as for AllSP\_DATA, but for the land bird dataset versions (excluding the the marine, coastal, wetland and riverine species). Stored as a list with six elements:

* 1-6: The same as for AllSP\_DATA, but with the land bird files. Note that there are two fewer datasets (i.e. n = 48 in total [49 when the embargo is lifted]; n = 8 for the alien datasets, and n = 6 for the historic period datasets). There were no prehistoric period analyses undertaken using the land bird datasets.

1. TRAIT\_DATA – a list with two elements:

* The trait data for extant species (BirdTree taxonomy). Stored as a data frame with 9,993 rows and 12 columns (the 9 traits in the analyses and family, order and habitat data). Rows are species and columns trait information. Data are from AVONET (https://onlinelibrary.wiley.com/doi/full/10.1111/ele.13898) and users should cite this paper if using these data in any publication (it is probably preferable to source the data direct from AVONET in this case). See AVONET metadata for information on all traits.
* The trait data for extinct species. A data frame with 159 rows and 15 columns. Rows are species and columns trait information. Contains data for the same 9 traits as for the extant species, in addition to the Archipelago the species was found on, the Order, Family and Genus, as well as whether or not the species is Aquatic (TRUE = aquatic). Note that some trait measurements are from museum skins, while others are gaps that have been imputed / estimated.

1. CONS\_TREE – contains the consensus phylogeny (a ‘phylo’ object), with 10,154 tips. Consensus tree built using multiple Jetz et al. (2012) BirdTree trees.
2. ETP – contains the elevation (Elev\_max), climate (Bio1\_m, Bio12\_m), isolation (Iso), MeanDist and island type (Type\_) information for all datasets.