Meta data for Latitudinal and anthropogenic effects on the structuring of networks linking blood-feeding flies and their vertebrate hosts

Ben Bellekom, Talya D. Hackett and Owen T Lewis.

**Overview**

These data consist of records of interactions between biting Diptera and their host species, extracted from literature.

In the file Latitudinal and anthropogenic full data.csv, interactions are grouped by the manuscript they were extracted from (WebID column). The article and author names are given for each web. Taxonomic information and the frequency are recorded for each interaction pair. Each web is assigned to one of three habitat classifications based on descriptions from the manuscripts, or where sufficient description is absent, we classified the site using the site coordinates and satellite imagery (QGIS in combination with Google Earth) to infer the habitat.

In the file Site information.csv, site and basic web summary statistic are recorded for each network.

The code used in this analysis is found in the file Latitudinal and anthropogenic Script.R

N.B. Village/Urban habitats are referred to as Urban in the datasets and R script.

**Metadata**

The file Latitudinal and anthropogenic full data.csv contains 16 columns

diptera\_family – identifies the Diptera family

diptera\_genus – identifies the Diptera genus

diptera\_species- identifies the Diptera species

host\_species- identifies the host species

host\_genus- identifies the host genus

host\_family- identifies the host family

host\_order – identifies the host order

host\_class- identifies the host class

webID- identifies the manuscript the data were extracted from and is used an as identify in the Bipartite package

frequency- is the total number of interactions by the species pair in the web

country- country where the data were collected

latitude – site latitude

longitude – site longitude

Site- habitat classification (Agricultural, Near-natural, Urban)

Paper name – title of the manuscript from which the data were extracted

Authors – authors of the manuscript from which the data were extracted

The file Site information.csv contains 6 columns

Name- identifies the webID, consistent with the previously described file

Site- habitat classification (Agricultural, Near-natural, Urban)

Latitude- site latitude

Longitude – site longitude

Numhost- the number of host species in the web

Numinsect- the number of Diptera species in the web

Matrixsize- the total number of interactions in the web

**Data analysis**

Data were analysed in R primarily using Generalised linear models (GLM). All analyses are described in the manuscript and annotations within the R script (file: Latitudinal and anthropogenic Script.R) identify and describe each analysis step. The packages bipartite, *iNEXT, tidyverse, lme4* are required for the primary analysis and *ggplot2, ggpubr, iNEXT* and *bipartite* for data visualisation.