

## ARTICLE

## Functional traits predict species co-occurrence patterns in a North American Odonata metacommunity

Francesco Cerini<sup>1,2</sup>  | Leonardo Vignoli<sup>3</sup>  | Michael Blust<sup>4</sup> | Giovanni Strona<sup>5,6</sup> <sup>1</sup>Dipartimento Scienze Ecologiche e Biologiche, Università della Tuscia, Viterbo, Italy<sup>2</sup>School of Biological Sciences, University of Bristol, Bristol, UK<sup>3</sup>Department of Science, University of Roma Tre, Rome, Italy<sup>4</sup>Department of Biology, Green Mountain College, Poultney, Vermont, USA<sup>5</sup>European Commission, Joint Research Centre, Ispra, Italy<sup>6</sup>Faculty of Biological and Environmental Sciences, Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland

## Correspondence

Giovanni Strona

Email: [giovanni.strona@ec.europa.eu](mailto:giovanni.strona@ec.europa.eu)

Handling Editor: Robert R. Parmenter

## Abstract

The probability of occurrence of a given species in a target locality and assemblage is conditioned not only by environmental/climatic variables but also by the presence of other species (i.e., species co-occurrence). This framework, already complex in nature, becomes even more complicated if one considers the functional traits of species that, in turn, might influence the structure of metacommunities in various ways. Depending on the ecological and environmental setting, functional similarity (i.e., convergence in morphological and ecological traits) between species might either reduce their co-occurrence due to high niche overlap driving negative interactions or promote it if the similar traits are associated with local habitat suitability. Similarly, functional divergence might either promote species co-occurrence by limiting negative interactions through niche separation or reduce it through trait mediated environmental filtering. Therefore, discriminating between these alternative scenarios—predicting whether two species will tend to co-occur or not based on their traits—is extremely challenging. Here, we develop a novel protocol to tackle the challenge, and we demonstrate its effectiveness by showing that ecological species traits can predict species co-occurrence in a large dataset of North American Odonata. To this end, we first used the Hierarchical Modeling of Species Communities framework to quantify the pairwise species co-occurrence after controlling for environmental and climatic factors. Then, we used machine learning to generate models which proved capable of predict accurately the observed co-occurrence patterns from species functional traits. Our approach offers a generalizable analytical framework with the potential to clarify long-standing ecological questions.

## KEYWORDS

assembly rules, co-occurrence, HMSC, machine learning, Odonata, species interactions, species traits

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Ecosphere* published by Wiley Periodicals LLC on behalf of The Ecological Society of America.

## INTRODUCTION

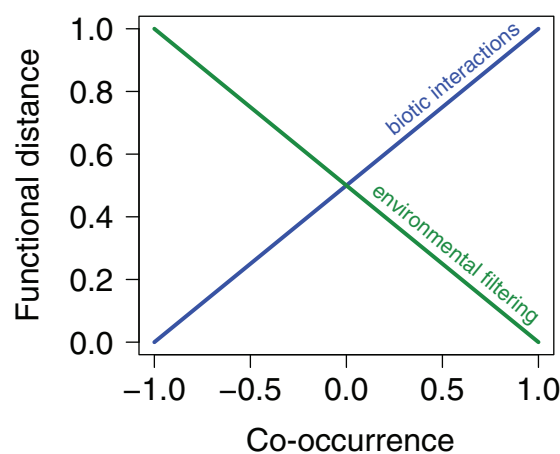
In community ecology, the distribution of the species in assemblages (i.e., metacommunities) is generally explained by three major ecological processes and their interactions: dispersal constraints, environmental (or habitat) filtering, and biotic interactions (Cadotte & Tucker, 2017; Hardy et al., 2012). While investigating the potential role of the latter in structuring bird communities in Melanesia, Diamond (1975) suggested that a high frequency of checkerboards (i.e., avoidance patterns in species  $\times$  sites 1/0 matrices) would emerge from interspecific competition due high similarity in the ecological niche of the species (e.g., similar resource use). This idea—and the statistical approaches which led Diamond to conclude that the observed frequency of checkerboards diverged substantially from a null expectation—promoted a famous controversy which is yet to be settled (Connor & Simberloff, 1979; Connor et al., 2013; Diamond et al., 2015; Strona et al., 2018).

The issue has received renewed interest in the context of ecological network research, where the question of whether checkerboard patterns indicate competition has been translated into whether interaction links can be inferred from species co-occurrence (Blanchet et al., 2020; Freilich et al., 2018; Morales-Castilla et al., 2015; Morueta-Holme et al., 2016), and with the development of joint species distribution models (jSDMs). In jSDMs, the probability of occurrence of a given species in a target locality is assessed not only based on environmental/climatic variables (as in standard SDMs), but also conditionally on the presence of other species (Ovaskainen et al., 2017; Tikhonov et al., 2020). One of the outputs of jSDMs is the residual co-occurrence of species after controlling for the contribution of abiotic variables and other confounding covariates, thus potentially accounting for the role of environmental filtering in the distribution patterns. It has been proposed that such residual co-occurrence could reflect ecological associations (Burner et al., 2021; Ovaskainen & Abrego, 2020; Ovaskainen et al., 2017) but, as for Diamond's work on assembly rules, the idea has received various criticisms (Clark et al., 2014; Dormann et al., 2018).

The main caveat is that even when accounting for many environmental variables, it is virtually impossible to fully control for the effect of environmental filtering that acts also at the microhabitat scale. Hence, one cannot rule out the possibility that the observed degree of co-occurrence between species is more driven by overlap (or segregation) in some dimensions of their ecological niches not included in the model rather than by direct ecological interactions (Barner et al., 2018; Blanchet et al., 2020). When we take into consideration species'

functional traits, we bring new potential information, but additional complications in the framework. Species traits play simultaneously fundamental roles in how a species interact with its habitat (environmental filtering) and with other species (biotic interactions; McGill et al., 2006). In fact, interspecific trait similarity can promote either species positive co-occurrence through consistent environmental filtering or negative co-occurrence in case the traits overlap promote species' competition for resources (Kohli et al., 2018). Likewise, trait dissimilarity might promote not only positive co-occurrence by reducing niche overlap and enhancing species' coexistence, but also negative co-occurrence through habitat filtering (i.e., species are best suited for living in separate locations).

In a simplified view, trait distance and co-occurrence should display a negative relationship if the traits drive primarily environmental filtering, and a positive one if the traits drive patterns of competitive exclusion (Figure 1). Clearly, in the real world (and considering multiple traits), the two processes coexist, making it hard to find a clear association. For instance, Elo et al. (2021) found no relationship between functional species similarity and negative co-occurrence on a large set of stream macroinvertebrate communities. Similarly, Burner et al. (2021) found that the difference in species traits was not a consistent predictor of the spatial associations for forest beetles. Nevertheless, it is important to consider that competitive dynamics (and species interactions in general) might be linked to functional traits in a more complex fashion than that provided by trait similarity.



**FIGURE 1** Theoretical relationships between species distance in functional traits and degree of co-occurrence. In a situation where functional traits drive environmental filtering, we should expect a negative relationship between distance and co-occurrence (green line); in the opposite scenario, where functional traits drive patterns of competitive exclusion, we should expect a positive relationship between distance and co-occurrence (blue line).

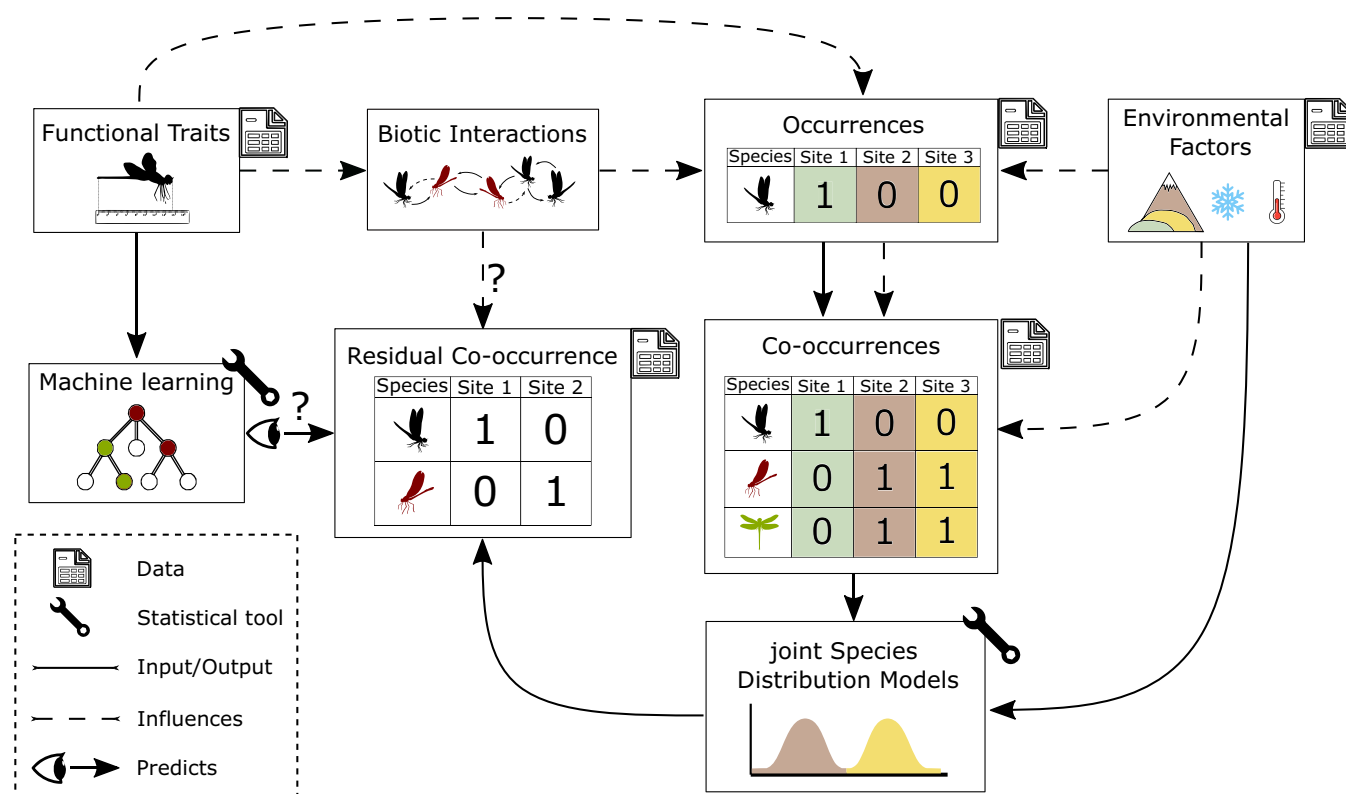
Nonlinear combinations of traits can either promote or weaken co-occurrence, hence making it difficult to identify signs of biotic interactions or microhabitat filters just by means of correlation with traits similarity. Here, by using a large dataset of dragonflies and damselflies (Odonata), we show that machine learning can harness such complexity, making it possible to accurately predict co-occurrence patterns from species functional traits (see Figure 2 for an overview of the conceptual framework).

## METHODS

### Dataset

We focused on a large dataset of Odonata from the state of Vermont in North America. Our choice was dictated by two reasons. First, Odonata are among the most well-known insects in terms of biology, ecology, and distribution, thanks to years of scientific research (Corbet, 1999) and amateur collections and observations that

contributed to build large datasets on species detailed distribution (e.g., “OdonataCentral,” Abbott, 2006–2022). Second, various studies have suggested that, in addition to habitat preference, interspecific interactions play a fundamental role in shaping Odonata communities and metacommunities (Cerini et al., 2019, 2020; Renner et al., 2019; Suutari et al., 2004). The data represent an updated and georeferenced version of the dataset from Blust and Pfeiffer (2015), a publication which summarized more than 80 years of Odonata collection activities (both scientific and amateur) in the state of Vermont. Each local site in the dataset consists of a water body (e.g., pond, brooks, lakes, and river beds) that was visited once or multiple times over the years. We cleaned the dataset by removing species with uncertain or missing taxonomical information. Presence data represent sampling of adults and/or larvae. The final dataset includes 4995 species occurrence records for 131 species across 917 sites and 86 years (1663 combinations of sites and years). We excluded all sites hosting only 1 species (i.e., not hosting an Odonata multispecies community),



**FIGURE 2** Diagram showing the logic behind our analytical framework. Species functional traits affect biotic interactions (e.g., competition for resources) and species habitat suitability (thus occurrences). Biotic interactions, together with climatic and environmental factors are major determinants of species distribution/occurrences. Co-occurrence patterns emerging from species occurrences are therefore simultaneously affected by biotic interactions and environmental drivers of species distribution, both mediated by functional traits. That is, species might tend to be found together (or not) because of overlap (or segregation) in their ecological niches and/or positive or negative interactions. Joint species distribution models (jSDMs) have been proposed as a potential statistical tool to distill the “residual” co-occurrence due to biotic interactions from that due to other (environmental) factors. Here we test if machine learning can use functional traits to model accurately pairwise positive and negative co-occurrence devised by jSDMs.

and then we excluded all species present in less than 10 sites to avoid considering very rare species sampled just sporadically throughout the years, which can potentially create artefactual negative co-occurrence pattern with many other species just due to their rarity. This left us with 520 sites, 764 combinations of sites/years, and 95 species (3712 occurrences).

For each site/year combination, we identified and obtained information on the most important abiotic factors known to shape Odonata communities (Bried & Siepielski, 2018) and to play a fundamental role in the phenology and the beginning of adult stage for Odonata (Corbet, 1999; Flenner et al., 2010), namely, temperature, precipitation, habitat type (lotic or lentic water body), elevation, and land cover. For temperature and precipitation, we matched the exact dates of the records (month/year) with actual temperature and precipitation data from the Global Historical Climatology Network monthly (GHCNm, freely available at <https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-monthly>) dataset, which provides monthly climate summaries from meteorological stations. We interpolated the temperature and precipitation values using ordinary kriging between the closest stations for which data were available for a target point. For this, we started from the 10 closest stations to the target point, and then we progressively increased the number of stations until we found at least three temperature and precipitation records. As for habitat type, we used information from the original data sources to classify sites into four broad categories, namely, lentic water, lotic water, wetland, and land. Additionally, we used a high-resolution dataset from the University of Vermont Spatial Analysis Laboratory (freely available at <https://geodata.vermont.gov/pages/land-cover>) to characterize the land cover surrounding the target localities. The dataset provides data for 2016 at a resolution of 0.5 m. Instead of focusing on the exact land cover class at each sampling point, we computed the fraction of area covered, respectively, by tree canopy, grassland/shrubland, bare soil, water, or artificial structures within a square of 500 × 500 m centered on the target site. Such a measure offered the most informative picture of the environmental setting available for the sampling localities, with the assumption that the landcover surrounding the sites stayed stable during the years.

As for the species traits, we selected nine morphology-, behavior-, and habitat use-related Odonata species traits (Appendix S1: Table S1) from the online free available Odonate Phenotypic Database (<http://www.odonatephenotypicdatabase.org>). These traits are measured on the adult stage of the species. Among the possible 33 traits available in the database, we excluded macro-geographical and location related traits (e.g., all

Odonata species of the study share the same continent, ecozone, and macroclimate), and we retained the nine traits that had the higher coverage across the study species. The information on the remaining traits for the study species were mostly missing.

## Statistical analyses

To assess pairwise species co-occurrence, we fit a latent-variable jSDM using the Hierarchical Modeling of Species Communities (HMSC) Bayesian framework. We fitted the models with a probit link function under the default prior distribution. We devised the following model (Equation 1):

$$\begin{aligned} \text{sp\_occ} = & \text{habitat} + \text{elevation} + \text{precipitation}^2 \\ & + \text{temperature}^2 + \text{tree canopy fraction} \\ & + \text{shrubland/grassland fraction} \\ & + \text{bare soil fraction} + \text{water fraction} \\ & + \text{artificial land fraction.} \end{aligned} \quad (1)$$

We included sites and years as random effects term, and we sampled the posterior distribution using four Markov chain Monte Carlo (MCMC) chains including 1000 samples each, with a thinning of 100 (for a total of 100k iterations), hence a burn-in of 25,000 iterations. We evaluated MCMC convergence by examining the distributions of the potential scale reduction factor over the parameters related to the fixed effects and the random effects (Gelman & Shalizi, 2013). We assessed model discriminatory power via AUC (area under the receiver operating characteristic curve, Pearce & Ferrier, 2000) and Tjur's  $R^2$  (2009).

We first explored the potential relationship between species trait dissimilarity and co-occurrence (i.e., theoretical framework in Figure 1). For each species pair, we assessed trait dissimilarity as the Mahalanobis distance (De Maesschalck et al., 2000) between their functional traits (using one-hot encoding of categorical variables with more levels), and we regressed such distances against the corresponding pairwise residual co-occurrence values derived from the HMSC analysis. We also explored the frequencies of correlations between pairwise species co-occurrence and the corresponding distances between individual traits (since most considered traits were categorical, such distances usually translated into 0/1 values, i.e., 0 if the target trait was shared by the two species, and 1 otherwise).

Then, we tested three different models aimed at predicting co-occurrence patterns from species functional traits: model I, a Random Forest regressor aimed at modeling the expected sign and intensity of the co-occurrence from the HMSC between any two species

in the dataset; model II, a Random Forest classifier aimed at predicting whether two species will have a positive ( $>0.5$ ) co-occurrence; model III, a Random Forest classifier aimed at predicting whether two species will have a negative ( $<0.5$ ) co-occurrence.

For all Random Forest models (Breiman, 2001), we first performed a hyper-parameter tuning to optimize model settings. For the models II and III, we performed 100 replicates by resampling the complete dataset to balance presence and absences. We also replicated the models using different thresholds for co-occurrence (0.4, 0.45, 0.50, 0.55, ..., 0.9) and then explored how the choice of the threshold affected model accuracy.

The three models have the following form (Equation 2):

$$\text{cooc}_{ij} \sim t1_i + t1_j + t2_i + t2_j + \dots + tn_i + tn_j, \quad (2)$$

where  $\text{cooc}_{ij}$  is the residual co-occurrence value derived from the HMSC analysis, and each entry  $tx_i$  and  $tx_j$  represents the  $x$ th trait of species  $i$  and  $j$  respectively (Appendix S1: Table S1).

For model I,  $\text{cooc}_{ij}$  corresponded to the continuous mean co-occurrence values from the HMSC analysis (theoretically ranging from  $-1$  to  $1$ ). For model II, we set  $\text{cooc}_{ij}$  to 1 for the pairs exhibiting a strong positive co-occurrence (i.e., mean co-occurrence  $>0.5$ ), and to 0 to the others. For model III, we set  $\text{cooc}_{ij}$  to 1 for the pairs exhibiting a strong negative co-occurrence (i.e., mean co-occurrence  $<-0.5$ ), and to 0 to the others.

To evaluate the predictive power of the Random Forest regressor (model I), we performed 100 replicates where we trained a model on a random set including 80% of the observations, testing it on the remaining 20%. We quantified the overall goodness of fit as the squared Pearson correlation coefficient between all the observed and the predicted co-occurrence values across the 100 simulations. For the Random Forest classifiers (models II and III), we assessed model quality by looking at type 1 and 2 classification errors and the resulting model accuracy based on internal out of bag validation (Janitza & Hornung, 2018; Mi et al., 2017). We performed the Random Forest analyses using the randomForest R package (Liaw & Wiener, 2002). All the code and data needed to replicate the analyses are available from Zenodo: <https://doi.org/10.5281/zenodo.10069033> (Strona, 2023).

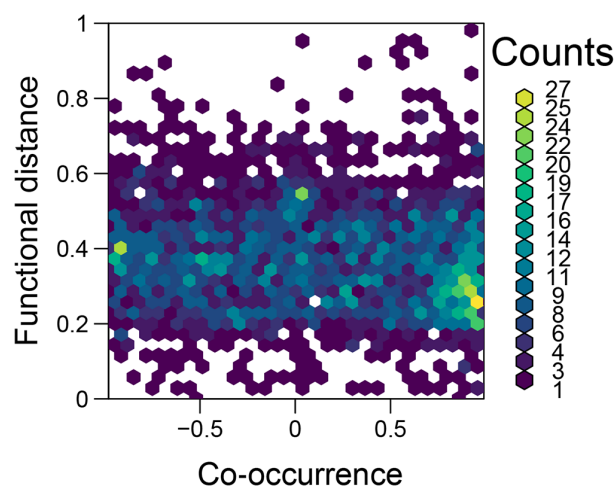
## RESULTS

The resulting HMSC model converged properly, with high effective sample sizes and potential scale reduction factors close to one, and showed an acceptable

performance (comparable with recent studies focusing on species' response to environmental and climatic change, such as Antão et al., 2022), with an average AUC of 0.92 and an average Tjur  $R^2$  of 0.17 (Appendix S1: Figure S1). On average, environmental and climatic variables accounted for 48.3% of the total explained variance, while the random year and site factors accounted, respectively, for 5.2% and 46.5% of the total explained variance. Among the environmental variables, land use (i.e., the combined contribution of the different fractions of land use categories surrounding a given site) resulted the most important factor (accounting for 20% of the total explained variance), followed by temperature (12.7%), elevation (5.4%), habitat (5.3%), and precipitation (4.9%) (Appendix S1: Figure S2). Among all possible species pairs ( $n = 4465$  for 95 species), 28.6% showed a positive co-occurrence ( $>0.5$ ), while 18.4% showed a negative co-occurrence ( $<-0.5$ ) (Appendix S1: Figure S3; we refer here to co-occurrence with a statistical support  $>0.95$ , see Methods).

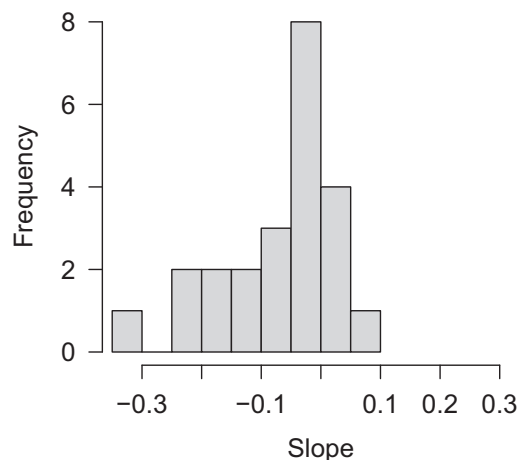
Overall, we found no relationship between species co-occurrence and trait dissimilarity ( $R^2 = 0.007$ ; Figure 3), as we recorded cases of high and low co-occurrence for both small and large functional distance (Figure 3). In most cases, however, functional distance fell into an average range. The correlations between individual trait distance and co-occurrence were very weak, but negative slopes were more common than positives (Figure 4).

Nevertheless, our machine learning approach revealed a significant potential for functional traits to predict co-occurrence patterns. Specifically, model I, that



**FIGURE 3** Density color-coded hexplot showing the regression ( $R^2 = 0.003$ ) between Hierarchical Modeling of Species Communities residuals co-occurrence values and species trait dissimilarity, calculated as the Mahalanobis distance, between all pair of species in the dataset.





**FIGURE 4** Frequency histograms of the slopes of the regressions between pairwise species co-occurrences and individual traits distance.

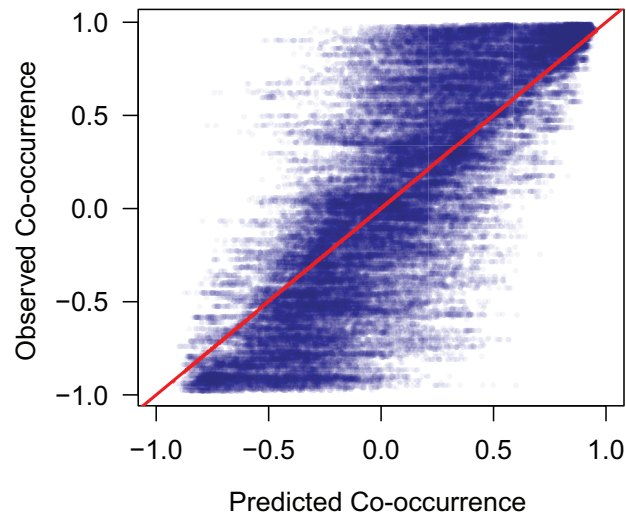
is, the Random Forest regressor aiming at predicting continuous positive and negative co-occurrence values, showed a  $R^2$  of 0.66 between all observed versus predicted co-occurrence values ( $n = 64,800$ ) obtained from 100 models trained on random sets including 80% of observations in the dataset and tested on the remaining 20% of observations (Figure 5).

The two models looking separately at positive (model II) and negative (model III) co-occurrence performed even better, with an average accuracy of 0.82 (0.008 SD) and 0.80 (0.009 SD, Figure 6). Error rates remained low regardless of the selected co-occurrence threshold (with type 2 errors always  $<0.2$  and type 1 errors always less than 0.3) and were lowest for thresholds around 0.5–0.6 (Figure 6). For a threshold of 0.5, model I had type 1 errors on average of 0.183 (0.011 SD) and 0.181 (0.009 SD), while model II had average type 1 and 2 errors of 0.240 (0.014 SD) and 0.154 (0.011 SD), respectively.

Among the considered traits, the adult phase morphological parameters related to body size emerged consistently as main predictors of co-occurrence (both positive and negative) (Figure 7). Specifically, the body length and hindwing length resulted by far as the major predictors, followed by the body color pattern and the type of aquatic habitat frequentation. Other traits related to flight behaviors and body pigmentation did not show strong effects (Figure 4; Appendix S1: Table S1).

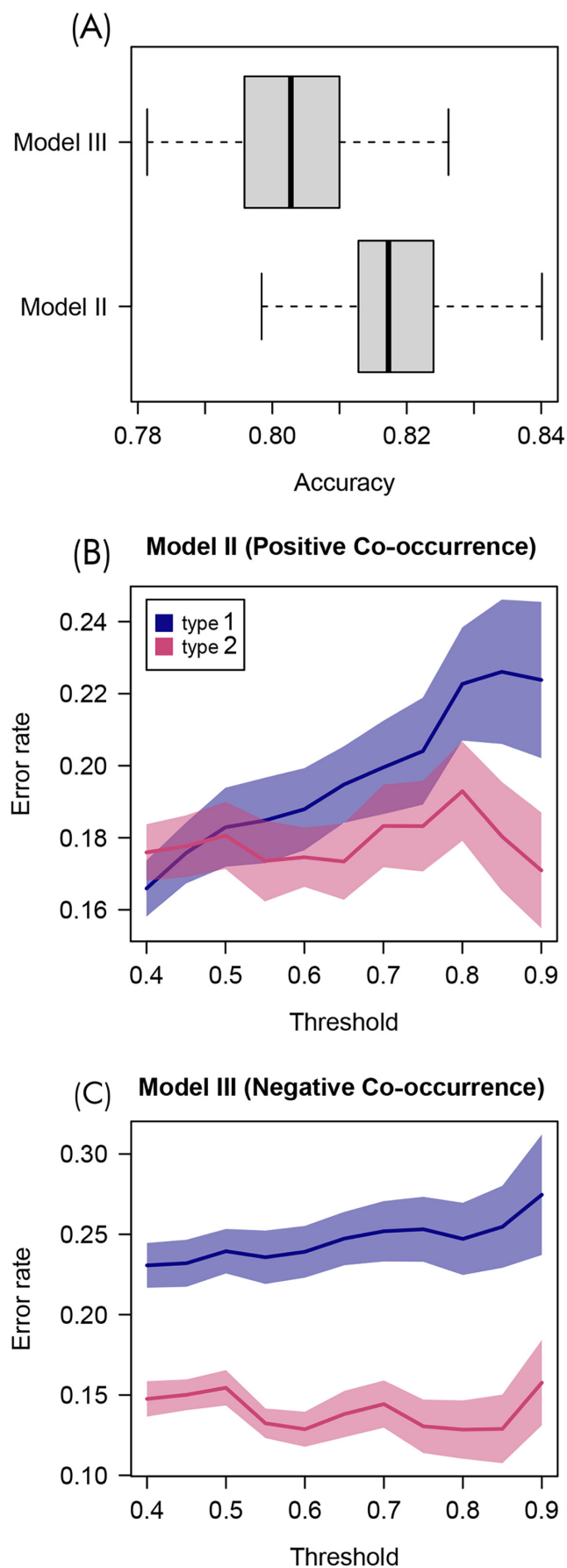
## DISCUSSION

Environmental features and interspecific interactions are assumed to be the main drivers of species distribution and abundance (Chesson, 2000), often interacting in a



**FIGURE 5** Observed residual co-occurrence from the Hierarchical Modeling of Species Communities analysis for all Odonata species pairs versus the predictions obtained from a Random Forest regressor modeling species co-occurrence based on species functional traits. The plot includes comparisons between observations and predictions for 100 models, each trained on a random sample including 80% of the available observations and tested on the remaining 20% (thus, each point in the plot correspond to test observations, not to training data). The red line is the line of equity.

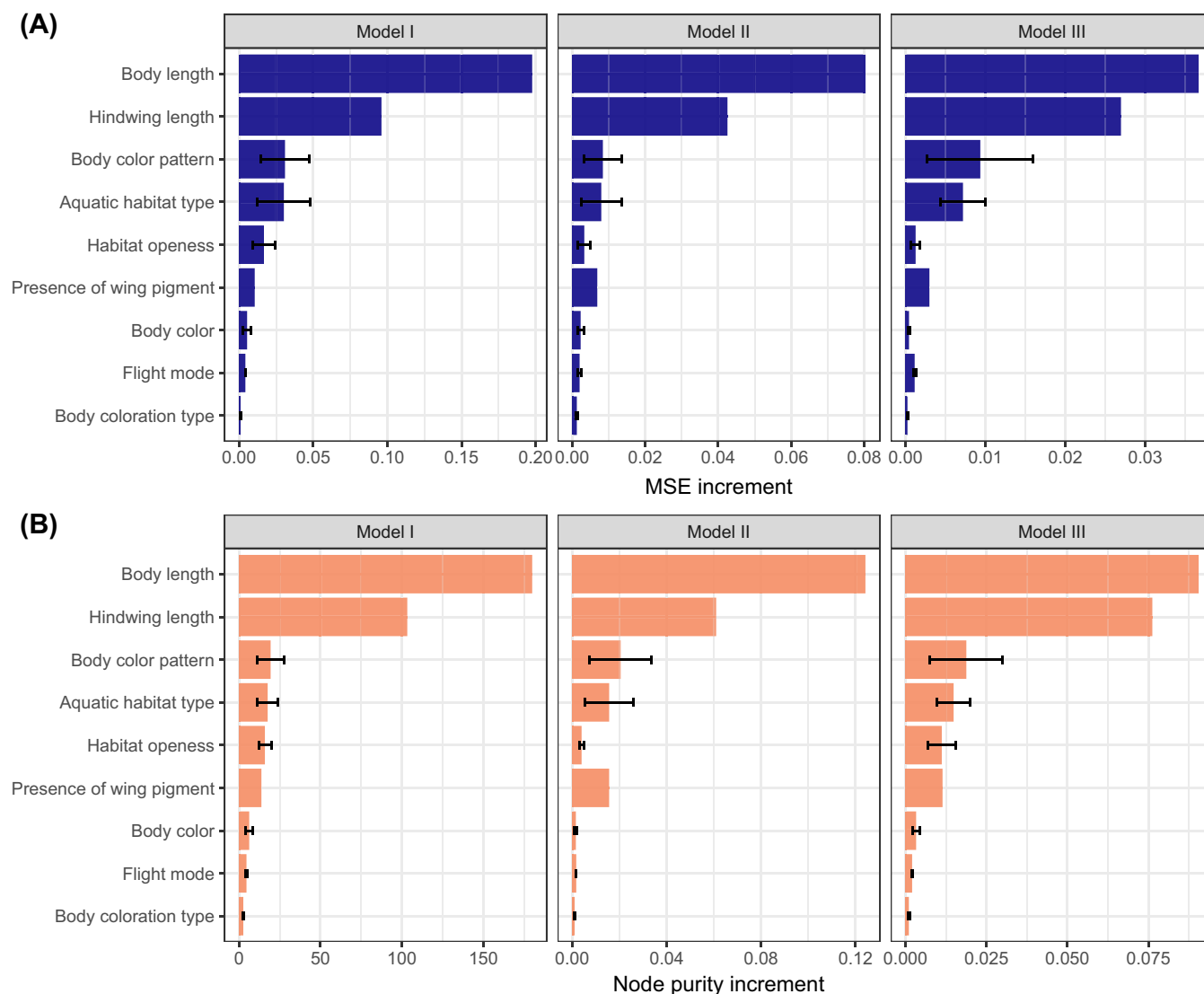
complex way in shaping communities composition (Ousterhout et al., 2015). For years, the difficulties in discerning the abiotic and biotic drivers of species occurrence and abundance patterns have fueled the still ongoing debate on the potential structuring role of assembly rules (Blanchet et al., 2020; Connor et al., 2013; Diamond et al., 2015; Gotelli & McCabe, 2002). Recently, new methodological frameworks have brought promising advances on the topic (Pollock et al., 2014; Tikhonov et al., 2020; Warton et al., 2015). Among these, jSDMs provide a valuable tool to disentangle the contribution that landscape features and habitat filtering (Mazerolle & Villard, 1999), rather than species interaction processes, have in structuring species occupancy and richness (D'Amen et al., 2018; Elo et al., 2021). Ideally, nonrandom co-occurrence patterns identified by jSDMs (such as the HMSC used here) are generated taking into account the weight that environmental variables have in explaining species presence-absence. Thus, in principle, such models should be able to disregard the co-occurrence patterns generated by habitat filtering processes acting at macroscale (e.g., landcover, average temperatures, and rainfall of the region) while retaining those emerging from interspecific interactions (Ovaskainen et al., 2017) or from uncontrolled microhabitat factors (e.g., water quality; Sun et al., 2021), if not



included in the model. Such an ability, which has been questioned in other works (Clark et al., 2014; Dormann et al., 2018), finds support in our results, which demonstrate how species functional traits can accurately predict co-occurrence patterns identified by HMSC.

Based on the theory of limiting similarity, species occurring syntopically with similar trait values should interact more strongly (i.e., resulting in lower fitness) than dissimilar species, thus leading to possible local exclusion (e.g., competitive checkerboards; Decaëns et al., 2008) and to niche specialization within the assemblages (MacArthur & Levins, 1967; Schirmer et al., 2020). Nevertheless, as observed in previous studies (Elo et al., 2021), our first analysis exploring the direct relationship between species traits dissimilarity and pairwise residual co-occurrence derived from the HMSC did not find any relevant pattern. In general, the correlations were very weak, but mostly negative, supporting a slightly more important role of environmental filtering over interactions (Figures 1 and 4). On the one hand, this result could be ascribed to a selection of habitat variables for the HMSC not accurately representing the environmental features acting as possible assembly filters (but see Bried & Siepielski, 2018; Cerini et al., 2021); thus, the co-occurrence patterns would still be an outcome produced by macro-environmental constraints not directly reflected in the species traits. Nonetheless, it is worth emphasizing that the environmental variables used in our model made up for a good portion of the explained variance, thus suggesting that the obtained residual co-occurrence might indeed be a measure not dependent on macroscale habitat filtering. On the other hand, the idea that similarities in species traits might reflect species competition for resources, and hence possibly the degree of co-occurrence (e.g., Elo et al., 2021), in a linear fashion is likely to be an oversimplified picture of the ecological reality, especially at a macroecological scale. Species traits affect the way species interact in complex and nonlinear ways that might not be captured by a distance measure. This is further complicated by the fact that while some trait combinations might be involved in

**FIGURE 6** Model quality assessment for the Random Forest classifiers (models II and III). (A) Model accuracy based on internal out-of-bag validation. (B) Type 1 and 2 error rates for model II (classifier modeling positive co-occurrence) versus different co-occurrence threshold selected to define a significant co-occurrence (e.g., a co-occurrence value >0.5, 0.6 etc., with a statistical support of 0.95 is defined as a 1, i.e., a “positive co-occurrence”). (C) Type 1 and 2 error rates for model III (classifier modeling negative co-occurrence) versus different co-occurrence threshold selected to define a significant co-occurrence.



**FIGURE 7** Importance of the species functional traits in the three Random Forest models to predict co-occurrence. (A) Traits importance calculated as the percentage increment of the mean square error (MSE) when permuting the out-of-bag data and extracting the resulting error increase when one variable is left out. (B) Traits important based on the random forest tree node purity, calculated as the reduction in sum of squared errors whenever a variable is chosen to split. Bars with whiskers show the mean scores  $\pm$  SEs of the different values for the one-hot encoded traits (i.e., traits with different categorical values like “habitat preference = open versus closed”). Model I: Random Forest and regressor modeling continuous positive and negative Hierarchical Modeling of Species Communities residual co-occurrence versus Odonata functional traits. Model II: Random Forest classifier modeling positive co-occurrence (i.e., co-occurrence  $>0.5$  with a statistical support of 0.95) versus Odonata functional traits. Model III: Random Forest classifier modeling negative co-occurrence (i.e., co-occurrence  $<-0.5$  with a statistical support of 0.95) versus Odonata functional traits. Detailed description of the species traits can be found in Appendix S1.

competitive dynamics and hence possibly promote negative co-occurrence, other trait combinations might be related, for instance, to habitat adaptation or density dependent processes (i.e., multispecies predation avoidance; Tirok & Gaedke, 2010) and thus favor positive co-occurrence, with this being especially true in freshwater ponds communities (Wellborn et al., 1996).

Here, we showed that machine learning permits to identify such complex and nonlinear links between

species functional traits and co-occurrence patterns. Although our main finding that species functional traits predict both negative and positive co-occurrence is not direct evidence (i.e., experimental) that interspecific interactions rather than microscale habitat filtering determine species distribution patterns, it provides novel and interesting baselines to deepen both hypothesized processes. The main assumption behind our approach is that functional traits are strong determinants of species



interactions and habitat suitability. Once controlled for the role of macroscale environmental variables in the HMSC-derived occurrence patterns, functional traits proved to predict the residual co-occurrences. This suggests that such species distribution patterns might retain at least some signature of species interactions. Focusing on the traits identified as most important predictors of negative co-occurrence by the model (Figure 7) supports such theoretical framework, as these are known to play important roles in interspecific competition and predation processes.

In both Anisoptera and Zygoptera, the body and hindwings lengths are good representation of species thermoregulatory capacity and flight potential (i.e., proxies for dispersal, competition, and predation capacities; Corbet, 1999; De Marco & Resende, 2002) and are easily linkable to negative interspecific interactions. However, we cannot abstract from our results the “sign” of the predictive effect of the traits on the occurrence—we do not know whether two species are predicted to segregate if their body length is more or less similar—thus, we have to discuss the ecological meaning of the two opposite cases. For example, the overlap in body size can indicate for a species pair the potential to use the habitat space in a similar way and directly compete as both adults (Boucenna et al., 2018) and larvae (e.g., same ecological guild, Suutari et al., 2004), in a way to generate local competitive exclusions and thus segregation patterns (Cerini et al., 2019). Indeed, Odonata revealed systematic morphological divergence between pairs of co-occurring species, especially those most demanding of environmental conditions (i.e., zygopterans, Oliveira-Junior et al., 2021). Conversely, a big difference in size could indicate the potential for the smallest species to be predated by the biggest, thus generating avoidance patterns (Cerini et al., 2019); this being especially true with Zygoptera–Anisoptera species pairs (Priyadarshana, 2021). In both cases, the importance of morphological traits in predicting the spatial pattern is a good indicator of how those traits can mediate the interaction processes potentially creating the negative co-occurrence.

Body and hindwings length were found to be good predictors also of positive co-occurrence (Figure 7). This might underlie cases where the functional traits relation with the species occurrence may not involve direct interactions but rather niche differences at microscale: for example, morphologically similar species living in the same sites by being dietary generalist or morphologically dissimilar species targeting preys with different size (Dudgeon, 1989). Similar-sized species that co-occur in the same sites might avoid the resource competition by being temporally segregated (e.g., partitioning the emergence timings; Boucenna et al., 2018), or by spatially

segregating within the local site (e.g., differential microhabitat use; Khelifa et al., 2013).

The Random Forest models predicted species co-occurrence from functional traits despite the many different generating processes, thus revealing a high potential of the proposed approach. Besides providing new hints for the long-standing debate on the ecological relevance of assembly rules (Diamond et al., 2015), this approach might produce interesting insights into the role of specific traits in determining species co-occurrence based on their predictive power in the different models. Generally speaking, as the species traits libraries keep being enriched (Jeliaskov et al., 2020), our framework could serve as an exploratory tool to identify which traits are potentially important determinants of co-occurrence, without requiring a priori knowledge of the exact mechanisms by which those functional traits affect interactions. The traits importance output might serve exactly as a basis to start formulating ecological/biological hypotheses to be tested thorough ad hoc experimental designs. The derived information from these theoretical and experimental approaches might shed light on the ecological meaning of those traits, possibly fueling investigation of phylogenetic signals of selection (Burns & Strauss, 2012). In conclusion, using machine learning to predict both positive and negative species co-occurrence patterns based on species functional traits appears as a promising tool with the potential to advance our understanding of whether, to what extent, and how species interactions and microscale environmental filters shape biodiversity patterns.

## AUTHOR CONTRIBUTIONS

Leonardo Vignoli, Giovanni Strona, and Francesco Cerini developed the idea together. Francesco Cerini and Leonardo Vignoli are to be considered as co-leading authors. Michael Blust provided the data. Giovanni Strona designed the statistical workflow and performed the analyses. Giovanni Strona, Francesco Cerini, and Leonardo Vignoli contributed equally to the first draft realization. All authors contributed to the final draft.

## CONFLICT OF INTEREST STATEMENT


The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Data and code (Strona, 2023) are available from Zenodo: <https://doi.org/10.5281/zenodo.10069033>.

## ORCID

Francesco Cerini  <https://orcid.org/0000-0002-1367-1239>

Leonardo Vignoli  <https://orcid.org/0000-0002-1599-092X>

Giovanni Strona  <https://orcid.org/0000-0003-2294-4013>

## REFERENCES

- Abbott, J. C. 2006–2022. “OdonataCentral: An Online Resource for the Distribution and Identification of Odonata.” <https://www.odonatacentral.org/#/>.
- Antão, L. H., B. Weigel, G. Strona, M. Hällfors, E. Kaarlejärvi, T. Dallas, Ø. H. Opedal, et al. 2022. “Climate Change Reshuffles Northern Species within Their Niches.” *Nature Climate Change* 12: 587–592.
- Barner, A. K., K. E. Coblenz, S. D. Hacker, and B. A. Menge. 2018. “Fundamental Contradictions among Observational and Experimental Estimates of Non-Trophic Species Interactions.” *Ecology* 99: 557–566.
- Blanchet, F. G., K. Cazelles, and D. Gravel. 2020. “Co-Occurrence Is Not Evidence of Ecological Interactions.” *Ecology Letters* 23: 1050–63.
- Blust, M., and B. Pfeiffer. 2015. “The Odonata of Vermont.” *Bulletin of American Odonatology* 11: 69–119.
- Boucenna, N., A. Kahalerras, N. Boukhemza-Zemmouri, M. Houhamdi, and R. Khelifa. 2018. “Niche Partitioning at Emergence of Two Sympatric Top-Predator Dragonflies, *Anax imperator* and *A. parthenope* (Odonata: Aeshnidae).” *Annales de la Société entomologique de France (N.S.)* 54: 73–80.
- Breiman, L. 2001. “Random Forests.” *Machine Learning* 45: 5–32.
- Bried, J. T., and A. M. Siepielski. 2018. “Opportunistic Data Reveal Widespread Species Turnover in Enallagma Damselflies at Biogeographical Scales.” *Ecography* 41: 958–970.
- Burner, R. C., J. G. Stephan, L. Drag, T. Birkemoe, J. Muller, T. Snäll, O. Ovaskainen, et al. 2021. “Traits Mediate Niches and Co-Occurrences of Forest Beetles in Ways that Differ among Bioclimatic Regions.” *Journal of Biogeography* 48: 3145–57.
- Burns, J. H., and S. Y. Strauss. 2012. “Effects of Competition on Phylogenetic Signal and Phenotypic Plasticity in Plant Functional Traits.” *Ecology* 93: S126–S137.
- Cadotte, M. W., and C. M. Tucker. 2017. “Should Environmental Filtering be Abandoned?” *Trends in Ecology & Evolution* 32: 429–437.
- Cerini, F., M. A. Bologna, and L. Vignoli. 2019. “Dragonflies Community Assembly in Artificial Habitats: Glimpses from Field and Manipulative Experiments.” *PLoS One* 14: e0214127.
- Cerini, F., P. Bombi, R. Cannings, and L. Vignoli. 2021. “Odonata Metacommunity Structure in Northern Ecosystems Is Driven by Temperature and Latitude.” *Insect Conservation and Diversity* 14: 675–685.
- Cerini, F., L. Stellati, and L. Vignoli. 2020. “Segregation Structure in Odonata Assemblages Follows the Latitudinal Gradient.” *Oecologia* 194: 15–25.
- Chesson, P. 2000. “Mechanisms of Maintenance of Species Diversity.” *Annual Review of Ecology and Systematics* 31: 343–366.
- Clark, J. S., A. E. Gelfand, C. W. Woodall, and K. Zhu. 2014. “More than the Sum of the Parts: Forest Climate Response from Joint Species Distribution Models.” *Ecological Applications* 24: 990–99.
- Connor, E. F., M. D. Collins, and D. Simberloff. 2013. “The Checkered History of Checkerboard Distributions.” *Ecology* 94: 2403–14.
- Connor, E. F., and D. Simberloff. 1979. “The Assembly of Species Communities: Chance or Competition?” *Ecology* 60: 1132–40.
- Corbet, P. S. 1999. *Dragonflies: Behaviour and Ecology of Odonata*. Colchester: Harley Books.
- D’Amen, M., H. K. Mod, N. J. Gotelli, and A. Guisan. 2018. “Disentangling Biotic Interactions, Environmental Filters, and Dispersal Limitation as Drivers of Species Co-Occurrence.” *Ecography* 41: 1233–44.
- De Maesschalck, R., D. Jouan-Rimbaud, and D. L. Massart. 2000. “The Mahalanobis Distance.” *Chemometrics and Intelligent Laboratory Systems* 50: 1–18.
- De Marco, P., Jr., and D. C. Resende. 2002. “Activity Patterns and Thermoregulation in a Tropical Dragonfly Assemblage.” *Odonatologica* 31: 129–138.
- Decaëns, T., P. Margerie, M. Aubert, M. Hedde, and F. Bureau. 2008. “Assembly Rules within Earthworm Communities in North-Western France—A Regional Analysis.” *Applied Soil Ecology* 39: 321–335.
- Diamond, J. M. 1975. “Assembly of Species Communities.” In *Ecology and Evolution of Communities*, edited by M. L. Cody and J. M. Diamond, 342–444. Cambridge: Harvard University Press.
- Diamond, J., S. L. Pimm, and J. G. Sanderson. 2015. “The Checkered History of Checkerboard Distributions: Comment.” *Ecology* 96: 3386–88.
- Dormann, C. F., M. Bobrowski, D. M. Dehling, D. J. Harris, F. Hartig, H. Lischke, M. D. Moretti, et al. 2018. “Biotic Interactions in Species Distribution Modelling: 10 Questions to Guide Interpretation and Avoid False Conclusions.” *Global Ecology and Biogeography* 27: 1004–16.
- Dudgeon, D. 1989. “Resource Partitioning among Odonata (Insecta: Anisoptera and Zygoptera) Larvae in a Hong Kong Forest Stream.” *Journal of Zoology* 217: 381–402.
- Elo, M., J. Jyrkänkallio-Mikkola, O. Ovaskainen, J. Soininen, K. T. Tolonen, and J. Heino. 2021. “Does Trait-Based Joint Species Distribution Modelling Reveal the Signature of Competition in Stream Macroinvertebrate Communities?” *Journal of Animal Ecology* 90: 1276–87.
- Flenner, I., O. Richter, and F. Suhling. 2010. “Rising Temperature and Development in Dragonfly Populations at Different Latitudes.” *Freshwater Biology* 55: 397–410.
- Freilich, M. A., E. Wieters, B. R. Broitman, P. A. Marquet, and S. A. Navarrete. 2018. “Species Co-Occurrence Networks: Can They Reveal Trophic and Non-Trophic Interactions in Ecological Communities?” *Ecology* 99: 690–99.
- Gelman, A., and C. R. Shalizi. 2013. “Philosophy and the Practice of Bayesian Statistics.” *British Journal of Mathematical and Statistical Psychology* 66: 8–38.
- Gotelli, N. J., and D. J. McCabe. 2002. “Species Co-Occurrence: A Meta-Analysis of J. M. Diamond’s Assembly Rules Model.” *Ecology* 83: 2091–96.
- Hardy, O. J., P. Couteron, F. Munoz, B. R. Ramesh, and R. Péliissier. 2012. “Phylogenetic Turnover in Tropical Tree Communities: Impact of Environmental Filtering, Biogeography and Mesoclimatic Niche Conservatism.” *Global Ecology and Biogeography* 21: 1007–16.
- Janitz, S., and R. Hornung. 2018. “On the Overestimation of Random Forest’s Out-of-Bag Error.” *PLoS One* 13: e0201904.
- Jeliazkov, A., D. Mijatovic, S. Chantepie, N. Andrew, R. Arlettaz, L. Barbaro, N. Barsoum, et al. 2020. “A Global Database for Metacommunity Ecology, Integrating Species, Traits, Environment and Space.” *Scientific Data* 7: 6.

- Khelifa, R., R. Zebba, A. Moussaoui, A. Kahalerras, S. Bensouilah, and H. Mahdjoub. 2013. "Niche Partitioning in Three Sympatric Congeneric Species of Dragonfly, *Orthetrum chrysostigma*, *O. coerulescens anceps*, and *O. nitidinerve*: The Importance of Microhabitat." *Journal of Insect Science* 13: 1–17.
- Kohli, B. A., R. C. Terry, and R. J. Rowe. 2018. "A Trait-Based Framework for Discerning Drivers of Species Co-Occurrence across Heterogeneous Landscapes." *Ecography* 41: 1921–33.
- Liaw, A., and M. Wiener. 2002. "Classification and Regression by randomForest." *R News* 2(3): 18–22. <https://CRAN.R-project.org/doc/Rnews/>.
- MacArthur, R., and R. Levins. 1967. "The Limiting Similarity, Convergence, and Divergence of Coexisting Species." *The American Naturalist* 101: 377–385.
- Mazerolle, M. J., and M.-A. Villard. 1999. "Patch Characteristics and Landscape Context as Predictors of Species Presence and Abundance: A Review." *Ecoscience* 6: 117–124.
- McGill, B. J., B. J. Enquist, E. Weiher, and M. Westoby. 2006. "Rebuilding Community Ecology from Functional Traits." *Trends in Ecology & Evolution* 21: 178–185.
- Mi, C., F. Huettmann, Y. Guo, X. Han, and L. Wen. 2017. "Why Choose Random Forest to Predict Rare Species Distribution with Few Samples in Large Undersampled Areas? Three Asian Crane Species Models Provide Supporting Evidence." *PeerJ* 5: e2849.
- Morales-Castilla, I., M. G. Matias, D. Gravel, and M. B. Araújo. 2015. "Inferring Biotic Interactions from Proxies." *Trends in Ecology & Evolution* 30: 347–356.
- Moruela-Holme, N., B. Blonder, B. Sandel, B. J. McGill, R. K. Peet, J. E. Ott, C. Violle, B. J. Enquist, P. M. Jørgensen, and J.-C. Svenning. 2016. "A Network Approach for Inferring Species Associations from Co-Occurrence Data." *Ecography* 39: 1139–50.
- Oliveira-Junior, J. M. B., M. A. Teodósio, and L. Juen. 2021. "Patterns of Co-Occurrence and Body Size in Dragonflies and Damselflies (Insecta: Odonata) in Preserved and Altered Amazonian Streams." *Austral Entomology* 60: 436–450.
- Ousterhout, B. H., T. L. Anderson, D. L. Drake, W. E. Peterman, and R. D. Semlitsch. 2015. "Habitat Traits and Species Interactions Differentially Affect Abundance and Body Size in Pond-Breeding Amphibians." *Journal of Animal Ecology* 84: 914–924.
- Ovaskainen, O., and N. Abrego. 2020. *Joint Species Distribution Modelling: With Applications in R*. Cambridge: Cambridge University Press.
- Ovaskainen, O., G. Tikhonov, A. Norberg, F. Guillaume Blanchet, L. Duan, D. Dunson, T. Roslin, and N. Abrego. 2017. "How to Make More Out of Community Data? A Conceptual Framework and Its Implementation as Models and Software." *Ecology Letters* 20: 561–576.
- Pearce, J., and S. Ferrier. 2000. "An Evaluation of Alternative Algorithms for Fitting Species Distribution Models Using Logistic Regression." *Ecological Modelling* 128: 127–147.
- Pollock, L. J., R. Tingley, W. K. Morris, N. Golding, R. B. O'Hara, K. M. Parris, P. A. Vesik, and M. A. McCarthy. 2014. "Understanding Co-Occurrence by Modelling Species Simultaneously with a Joint Species Distribution Model (JSDM)." *Methods in Ecology and Evolution* 5: 397–406.
- Priyadarshana, T. S. 2021. "Do Predatory Adult Odonates Estimate Their Adult Prey Odonates' Body Size and Dispersal Ability to Proceed with a Successful Attack?" *Journal of Threatened Taxa* 13: 18949–52.
- Renner, S., E. Périco, M. S. Dalzochio, and G. Sahlén. 2019. "Ecoregions within the Brazilian Pampa Biome Reflected in Odonata Species Assemblies." *Austral Ecology* 44: 461–472.
- Schirmer, A., J. Hoffmann, J. A. Eccard, and M. Dammhahn. 2020. "My Niche: Individual Spatial Niche Specialization Affects within- and between-Species Interactions." *Proceedings of the Royal Society B: Biological Sciences* 287: 20192211.
- Strona, G. 2023. "giovannistrona/odonata\_cooccurrence: Code and Data Associated to the Paper 'Functional Traits Predict Species Co-Occurrence Patterns in a North American Odonata metacommunity' version: v1.0.0." Zenodo. <https://doi.org/10.5281/zenodo.10069033>.
- Strona, G., W. Ulrich, and N. J. Gotelli. 2018. "Bi-Dimensional Null Model Analysis of Presence-Absence Binary Matrices." *Ecology* 99(1): 103–115.
- Sun, Z., C. Zhao, W. Zhu, W. Zhu, J. Feng, S. Su, and T. Zhao. 2021. "Microhabitat Features Determine the Tadpole Diversity in Mountainous Streams." *Ecological Indicators* 126: 107647.
- Suutari, E., M. J. Rantala, J. Salmela, and J. Suhonen. 2004. "Intraguild Predation and Interference Competition on the Endangered Dragonfly *Aeshna viridis*." *Oecologia* 140: 135–39.
- Tikhonov, G., Ø. H. Opedal, N. Abrego, A. Lehtikainen, M. M. J. Jonge, J. Oksanen, and O. Ovaskainen. 2020. "Joint Species Distribution Modelling with the R-Package HMSC." *Methods in Ecology and Evolution* 11: 442–47.
- Tirok, K., and U. Gaedke. 2010. "Internally Driven Alternation of Functional Traits in a Multispecies Predator–Prey System." *Ecology* 91: 1748–62.
- Tjur, T. 2009. "Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination." *The American Statistician* 63: 366–372.
- Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. C. Hui. 2015. "So Many Variables: Joint Modeling in Community Ecology." *Trends in Ecology & Evolution* 30: 766–779.
- Wellborn, G. A., D. K. Skelly, and E. E. Werner. 1996. "Mechanisms Creating Community Structure across a Freshwater Habitat Gradient." *Annual Review of Ecology and Systematics* 27: 337–363.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Cerini, Francesco, Leonardo Vignoli, Michael Blust, and Giovanni Strona. 2023. "Functional Traits Predict Species Co-Occurrence Patterns In a North American Odonata Metacommunity." *Ecosphere* 14(12): e4732. <https://doi.org/10.1002/ecs2.4732>