

RESEARCH PAPER

Species-genetic diversity correlation in a metacommunity of urban pond invertebrates



Frank Johansson ^{a,*}, Yeserin Yildirim ^a, Chaz Hyseni ^{a,b}, Jani Heino ^c, Jacob Höglund ^a, Luis Mauricio Bini ^d

^a Department of Ecology and Genetics, Animal Ecology, Uppsala University, Norbyvägen 18D, Uppsala SE 75236, Sweden

^b Colossal Biosciences, Dallas 75247, TX, USA

^c Geography Research Unit, University of Oulu, Oulu, Finland

^d Departamento de Ecología, Universidade Federal de Goiás, Goiânia 74001-970, GO, Brazil

ARTICLE INFO

Keywords:

Species diversity
Genetic diversity
Urban ponds
Dispersal
Aquatic invertebrates
Conservation

ABSTRACT

Understanding patterns of species-genetic diversity correlations (SGDC) is important for conservation purposes because it allows us to infer whether conservation of species diversity (SD) influences conservation of genetic diversity (GD) and the other way around. Here, we studied SGDCs using aquatic macrofauna in a set of 31 urban ponds in the metropolitan area of Stockholm, Sweden. We also estimated how land use and pond environmental factors affect SD and GD. SD was estimated as species richness. GD was estimated in four focal species that differed in their dispersal abilities: *Asellus aquaticus* (Isopoda), *Haliphus ruficollis* (Coleoptera), *Planorbis planorbis* (Gastropoda), *Rana temporaria* (Amphibia), using double digest restriction associated DNA (ddRAD) sequencing data. There were no significant SGDCs for any of the species. Similarly, GD was not related to land use or pond environment. However, SD had a significant positive correlation with total invertebrate abundance and pond area. Given the absence of significant SGDCs in our study, and the mixed positive and negative patterns found in previous studies reporting SGDCs, we suggest that simultaneously preserving species and genetic diversity in urban areas will prove challenging.

Introduction

Biodiversity can be estimated at different levels, from genes to ecosystems (Bowen, 1999). Two common estimates are species diversity and genetic diversity. Species diversity (SD) is estimated as the number of species (i.e., species richness), or metrics such as H' (Shannon & Weaver, 1963) and J' (Pielou, 1966). Genetic diversity (GD) is measured by metrics such as number of segregating sites (S ; Nei, 1987), average number of nucleotide differences (K ; Tajima, 1983), or nucleotide diversity (π ; Nei, 1987). Both biodiversity estimates might be affected by the same factors, such as drift, migration, and landscape structure (Vellend, 2005). Hence, understanding species-genetic diversity correlations (SGDC) is fundamental from the perspective of theoretical evolutionary ecology (Vellend, 2003). It is also important from a conservation biology standpoint, because it helps practitioners understand whether conserving species diversity (SD) also protects genetic diversity (GD), and vice versa (Kahilainen et al., 2014). For example, maintaining high genetic diversity within populations may enhance their resilience to environmental changes and help support overall species diversity in

biotic communities.

GD is affected by population size, genetic drift, mutation, and migration (Vellend, 2005). SD is affected by migration, ecological drift, and environmental filtering, where the latter process refers to selection against certain species by the environment (Hubbell, 2001; MacArthur & Wilson, 1967). Since these variables might act in parallel on SD and GD, it has been suggested that a positive relationship should be the most common pattern observed (Vellend, 2003). For example, in its simplest form, a positive SGDC occurs when dispersal and environmental factors have similar effects on the genetic diversity of the focal species and the distribution of other species in the community (Vellend, 2003; Lamy et al., 2013). Conversely, negative SGDC is expected if dispersal and environmental factors have different effects on the genetic diversity of the focal species and the distribution of other species in the community. Interestingly, empirical studies have shown varying results, and there seems to be no consensus on whether SGDCs are positive or negative in general. In a review of this relationship, Lamy et al. (2017) found that 11% and 6% of the studies found significantly positive and negative correlations, respectively. For a better understanding, these authors

* Corresponding author.

suggested that empirical studies should decompose statistical analyses of SGDCs into environmental, population, and community factors (Lamy et al., 2017). Thus, it is important to explore how environmental variables are correlated with SD and GD, respectively, which may provide a more mechanistic understanding of the strength of SGDCs.

Urbanization has large impacts on biodiversity and structure of ecological communities (Dudgeon, 2006; Gál et al., 2019; Liu et al., 2022). Despite this, urban environments may harbor surprisingly high levels of biodiversity (Ives et al. 2016). For example, McKinney et al. (2011) found that bird species richness was higher in urban versus rural wetlands in Rhode Island, USA. However, the conjecture that urban ecosystems are more diverse than rural ecosystems is not a universal finding, and some studies have suggested the opposite for rural versus urban pond ecosystems (e.g., Hill et al., 2016). Ponds represent an important ecosystem type of urban aquatic environment harboring high biodiversity (Hill et al., 2016). Biodiversity in urban ponds has been found to be positively correlated with pond size, amount of vegetation in the pond, green area around the pond, macrophyte richness, and connectivity (Goertzen & Suhling, 2013; Heino et al., 2017; Hill et al., 2015; Thornhill et al., 2017). Most of these biodiversity studies in urban ponds have focused on species diversity, whereas we have very little information on biodiversity in terms of SGDCs.

Most SGDC studies have been conducted in rural areas (e.g., Ishii et al., 2022; Xie et al. 2013), whereas studies in urban areas are still lacking. It might therefore be interesting to explore SGDC patterns in urban environments. Biodiversity in urban areas tends to be concentrated within “green spaces.” Although these green spaces often harbor most of the urban biodiversity, they are highly fragmented (Angel et al., 2012), thus urban areas tend to be more fragmented compared to rural areas. Fragmentation usually results in a decrease of species and genetic diversity (Duncan et al., 2020; Fahrig, 2003; Reed & Frankham, 2003). This is especially true in scattered urban green spaces. These differences in fragmentation between urban and rural areas might result in a lower connectivity in urban areas compared to rural ones and thus lead to fewer dispersal opportunities for organisms in urban areas (Angel et al., 2012). Such differences have the potential to cause different patterns in SGDC between rural and urban areas. Currently the majority of SGDC studies have been conducted in rural areas (Lamy et al., 2017). To facilitate a robust comparison of SGDC between urban and rural areas, additional studies from urban areas are needed.

The main purpose of this study was to estimate SGDCs in ponds in an urban area. In general, SGDCs are assumed to be dependent on the degree of similarity of species niches and their dispersal abilities (Lamy et al., 2017). In this study, we limited our SGDC approach to dispersal abilities of selected species, and three important environmental pond variables assumed to be important for biodiversity in ponds. The main purpose of our study was to examine the relationship between species diversity and genetic diversity in a metacommunity of invertebrates in urban ponds. For genetic diversity, we focused on four species with different dispersal capacity. We did this because we wanted to examine how dispersal differences affect SGDCs. If the focal species and the rest of the species in the community have similar dispersal abilities a positive SGDC is predicted (Lamy et al., 2013). Thus, under the assumption that there is an average dispersal ability in a community we predict a positive SGDC in species with intermediate dispersal capabilities. For species with a low dispersal capacity, we predict a negative SGDC. A potential reason is that such species would not be able to reach some communities, but they might have high population densities in those communities where they thrive. For species with high dispersal, we predict positive SGDC, but a steeper slope compared to species with intermediate dispersal. Population densities and niche utilization could obviously modify these predictions, and there are numerous explanations for SGDC patterns. However, this study does not aim to cover all potential explanations. For further theoretical insights and detailed explanations, we refer readers to Lamy et al. (2017). In addition to exploring the effects of dispersal on the SGDC, we also explored how environmental,

spatial, and land use variables in and around ponds affected SD and GD.

Materials and methods

Species richness and abundance

In May-June 2019, we sampled aquatic invertebrates of four taxonomic insect orders, comprising Odonata (dragonflies; larvae), Trichoptera (caddisflies; larvae), Coleoptera (water beetles, both larvae and adults), and Hemiptera (semi-aquatic bugs, both larvae and adults), as well as the class Gastropoda (freshwater snails, adults) in 31 ponds (Fig. 1) in the city of Stockholm, Sweden. The city has ca. 900 000 residents, and when including the metropolitan area has ca. 1.5 million inhabitants. Ponds comprised both natural and man-made water bodies with an area ranging between 2 and 20,000 m² and holding water for at least 4 months of the year (Biggs et al., 2005).

Invertebrates were sampled with a hand net (diameter = 20 cm, mesh size = 1.5 mm). Six samples were taken in each pond at a depth of 20–30 cm. We performed 1 m sweeps with the net along the bottom in opposite directions (left to right) eight times, which constituted one sample. These six samples covered all types of microhabitats along the shoreline, e.g., soft and hard bottom, as well as with and without vegetation. The samples were preserved in 70% ethanol and the taxonomic identification was performed later in the laboratory. Our measure of species richness (SD) for each pond included taxa that could not be identified to the species level.

Local and land use variables

We sampled the following pond local pond variables: pond size (area), pond depth, pH, total organic carbon, total nitrogen, total phosphorus, percentage of floating vegetation in the pond, percentage of emerging vegetation in the pond, and the percentage of bushes surrounding the ponds. In addition, we estimated the land use variable: percentage of artificial surface around the ponds for a 250-m radius around each pond: (i.e., buildings, roads, sidewalks, parking lots, and other impervious surfaces). The methods for collecting these environmental variables have been described in detail in Hyseni et al. (2021), and below we provide a general description on how they were collected. The sampling and the data set is part of a larger study focusing on biodiversity in urban areas (Hyseni et al., 2021), and therefore some of the method description overlaps with previous studies.

Species-genetic diversity analysis (SGDC)

We used the following environmental variables in our analysis of SGDCs: pond size (area), percentage of floating vegetation in the pond, and percentage of artificial surface around the ponds (i.e., buildings, roads, sidewalks, parking lots, and other impervious surfaces). We chose these variables because they have been found to be significantly associated with species richness in other studies on biodiversity in urban ponds (Goertzen & Suhling, 2013; Heino et al., 2017; Hill et al., 2015; Thornhill et al., 2017). Additionally, we analyzed how the total abundance (defined as total number of individuals caught during sampling) of all invertebrates sampled in a pond was associated with genetic diversity. We added this variable because high invertebrate abundance is expected to lead to overall higher genetic diversity across taxa, but individual taxa such as our focal species may deviate from this expectation. Pond size was estimated from terrain maps available through Lantmäteriet (The Swedish mapping, cadastral and land registration authority) using QGIS (QGIS Development Team 2021). The percentage of floating vegetation cover was visually assessed using a categorical scale of 1–10, representing a range of 0 to 100% cover. Percentage of artificial surface around the ponds was calculated within a 250-m radius around each pond. Land cover data were retrieved from the Sentinel-2 Global Land Cover (<http://s2glc.cbk.waw.pl>). For more information

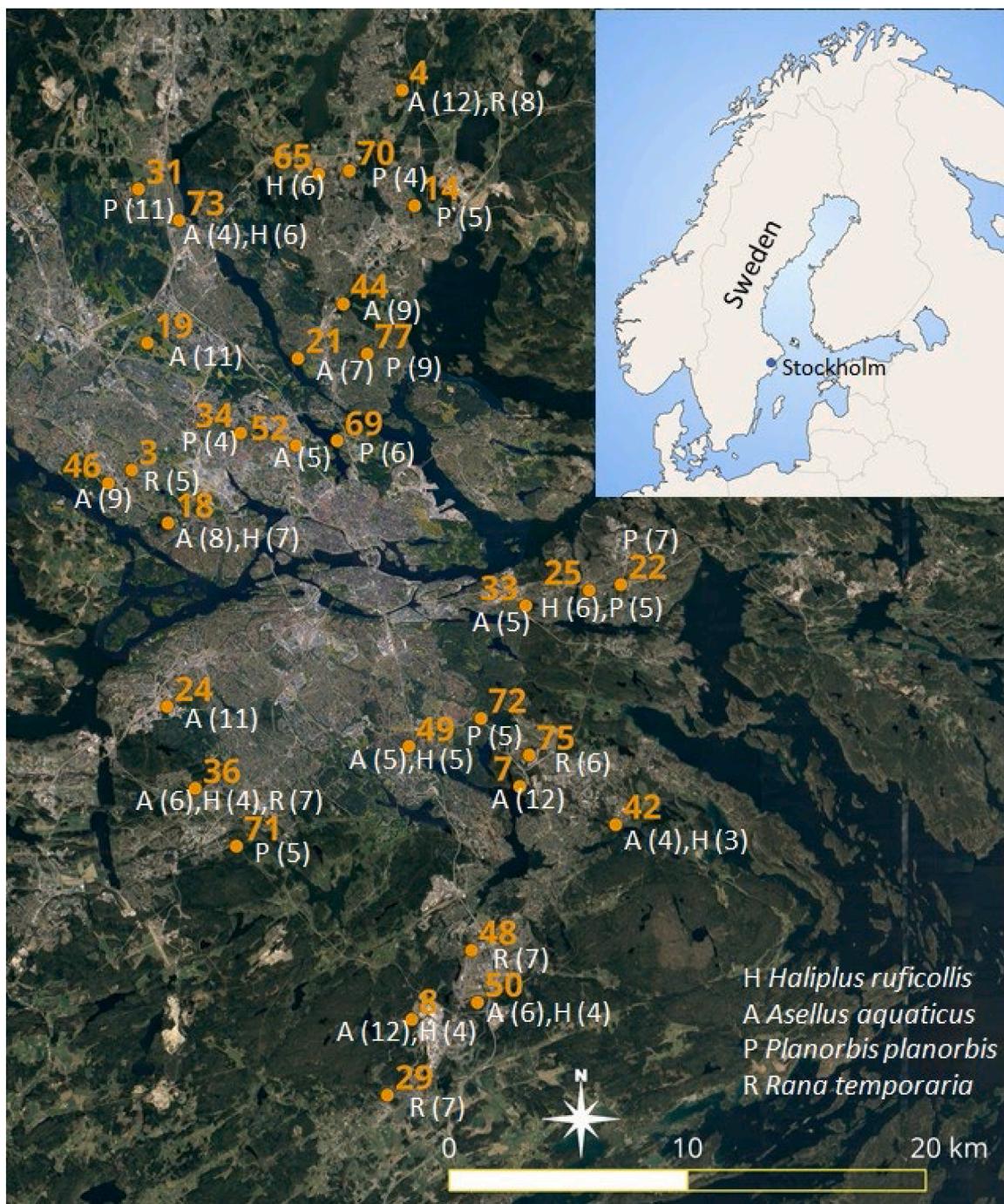


Fig. 1. Sampling locations for all four focal species. The yellow numbers are the codes for each location and the letters represent the species sampled. Gray color represents built up areas, dark green forest, light green open non-built-up areas, and dark blue water. The numbers in brackets next to each species code corresponds to the sample size for that specific location. The sampling pond locations in Stockholm were mapped in Quantum Geographic Information System (QGIS). The base map is available under non-restrictive creative commons license obtained from Wikimedia Commons, <https://commons.wikimedia.org/wiki/File:Scandinavia-template.png>.

on how these environmental variables were estimated and assembled see [Hyseni et al. \(2021\)](#).

Genetic analysis

For estimation of genetic diversity, we sampled three species of invertebrates in September 2020: *Asellus aquaticus* (Isopoda), *Haliplus ruficollis* (Coleoptera), *Planorbis planorbis* (Gastropoda), and one vertebrate species, *Rana temporaria* (Amphibia). The invertebrate species were selected for the analysis because they were widely distributed across numerous ponds, occurring in high abundance, and exhibited

varying dispersal capabilities. Amphibians were also included in the study due to their endangered status, and understanding the effects of genetic diversity in urban areas is crucial for conservation managers. The invertebrate taxa were sampled with sweep net samples along the shoreline for about 30 min in each pond.

Egg clutches of *R. temporaria* were collected along pond shores. Only a small fraction of eggs was gathered initially from each clutch. Upon arrival at the laboratory, one egg was selected from each clutch. The rest of the eggs were used for another purpose. Each clutch represents eggs from a single female. We were unable to collect samples of all species

from all of the ponds (see Table T1). We classified *H. ruficollis* as a high-dispersal species because this beetle has fully developed wings and can fly between aquatic habitats (Boda et al., 2014) and *R. temporaria* as a low-dispersal species because anurans generally have a low dispersal capacity (Sinsch, 1990). The other two taxa were classified as intermediate dispersers. We sampled *A. asellus* from 30 ponds ($n = 360$), *H. ruficollis* from 12 ponds ($n = 105$), *P. planorbis* from 13 ponds ($n = 126$) and *R. temporaria* from 8 ponds ($n = 66$). However, after double digest restriction-associated DNA (ddRAD) sequencing, we excluded some samples due to low DNA quality (see Table T1 for final sample numbers) and the remaining samples were used for further statistical analyses.

Library preparation, ddRAD-sequencing and calculation of genetic diversity

We made ddRAD sequencing libraries for 657 samples with a modified version of the protocols from the original ddRAD (Peterson et al., 2012) method and the methylation-sensitive version of ddRAD (Di Marsico et al., 2020; Marconi et al., 2019). While this reduces the number of loci available for this study, we did this for the purpose of also characterizing methylation levels in future studies. We extracted genomic DNA using the salting out method described by Till et al. (2015), with the following high-throughput modifications: extraction was performed in deep-well 2 mL 96-well plates, with volumes per sample of DNA binding buffer (6 M potassium iodide), lysis buffer, and 3 M sodium acetate reduced to 400 μ L, 500 μ L, and 150 μ L, respectively; lysis was performed overnight, and centrifugation was performed at reduced speeds (1500 $\times g$) for longer (10 min). Each DNA library was double-digested with both enzyme combinations (AciI + MseI and PstI + MseI), and a sample-specific barcoded adapter was ligated to the methylation-sensitive restriction end. A common Y adapter was ligated to the sticky end left by MseI. The process for the AciI enzyme, for example, was as follows: 1) restriction digestion: genomic DNA was treated with 5 units of AciI and MseI restriction enzymes (New England BioLabs), 2) adapter ligation: digested DNA was ligated to 0.2 μ M of unique barcoded adapters (barcodes by individual; see Appendix A: File 1) and 0.2 μ M of common Y adapter (MseI) using 1 unit of T4 DNA ligase (ThermoFisher), 0.2 mM ATP and 1x buffer for a final volume of 50 μ L. The libraries were then pooled as per the experimental design, purified and size selected by magnetic beads (Sera-Mag SpeedBeads; Cytiva) and E-gel electrophoresis (Invitrogen) for fragments in the range of 250 bp to 600 bp. Size-selected libraries were quantified using a fluorometer (Qubit; Life Technologies), and a normalized DNA amount (15 ng) was amplified with a primer that introduced an Illumina index (at the Y common adapter site) for demultiplexing by sampling site. Following PCR with uniquely indexed primers, multiple samples were pooled. PCR-enrichment was performed as described by Peterson et al. (2012). Specifically, 15 ng of pooled DNA was amplified using 1 U of Phusion High-Fidelity DNA polymerase (New England BioLabs) in a final volume of 50 μ L containing 1X Phusion HF Buffer, 0.2 μ M of AciI/PstI PCR primer, 0.2 μ M MseI index primer, and 0.2 mM dNTPs (0.05 mM each dNTP). Amplified libraries were purified with magnetic beads and then quantified (Qubit, Invitrogen; and Bioanalyzer, Agilent Technologies). The grouped libraries were pooled in an equimolar fashion, and the final library was Illumina-sequenced using 150-bp paired-end chemistry.

The libraries were sequenced in two lanes on an Illumina Novaseq 6000 from both directions (2×150 bp) at NGI (National Genomics Institute), SciLifeLab, Uppsala, Sweden. Trimming of adapters/primers and demultiplexing of the raw data were performed using CUTADAPT v4.0 (Martin, 2011) and fastq demux, respectively. The demultiplexed data was archived in the NCBI Sequence Read Archive BioProject ID: PRJNA1026545PRJN, link: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA1026545?reviewer=k77s47nm26he873ahrn69s8jslhttp>). All reads were trimmed to a uniform length of 100 bp, and reads with quality score < phred33 were discarded in trimmomatic v 0.39 (Bolger et al., 2014). The *de novo* pipeline in STACKS v.2.62 (Catchen et al., 2013) was used for SNP calling by running each step of the pipeline

(ustacks, cstacks, sstacks, tsv2bam and gstacks) separately. For the m parameter (i.e., coverage/stack depth), we used a value of three across all species, as this m value is suitable for most datasets (Rochette & Catchen, 2017). We performed parameter optimization for M (nucleotide differences allowed for stacks to be considered the same) and n (differences allowed across samples when building a population catalog of alleles and loci) in a subset of 16 randomly selected individuals for each species as described by Paris et al. (2017). We decided to use 4 for the M and n parameters for all species except *H. ruficollis* for which we used 3 (Appendix A: Fig. S1). We ran the *de novo* pipeline on all samples for each species separately using these parameters and other default settings. The catalog was further processed with the population unit in STACKS with an initial R60 setup, where a locus was retained if it was present in at least 60% of the individuals. We also chose a minimum minor allele frequency of 0.05, a maximum observed heterozygosity of 0.7 and one random SNP per locus. The populations step was repeated with several different R setups after removing samples with >50% missing data for each species. The aim was to find an R value that would yield the maximum number of individuals with an optimal number of loci. We used R75 for *A. aquaticus* and *H. ruficollis*, and R80 for *P. planorbis* and *R. temporaria*. After removing samples with more than a third of missing data, we then used the resulting dataset for all downstream analyses. The final data set contained 126 samples from 16 populations across 1336 loci for *A. aquaticus* (Appendix A: File 2). *Haliphus ruficollis* had 45 samples from 9 populations across 1179 loci (Appendix A: File 3). *Planorbis planorbis* had 61 samples, 10 populations, and 1922 loci (Appendix A: File 4). *R. temporaria* had 40 samples, 6 populations, and 1422 loci (Appendix A: File 5). Appendix A, Table T1 shows the number of individuals at each sampling location for each species.

To investigate SGDC, we used nucleotide diversity (π) and proportion of polymorphic nucleotide sites (PPL) as an estimate of genetic diversity. Arlequin 3.5 (Excoffier et al., 2005) was used for π , and the stacks populations module was used for PPL calculations. We decided to continue with only π as a Pearson correlation analysis showed a highly significant positive correlation between these two estimates for each species (data not shown).

Data analysis

We used linear mixed-effects modeling (Zuur et al., 2009) to test the relationship between nucleotide diversity (π) and the following explanatory variables (i.e., fixed effects): species richness, pond area, floating vegetation (%), percentage of artificial surface around ponds, and total abundance of invertebrates. In these models, the focal species (*H. ruficollis*, *P. planorbis*, *A. aquaticus* and *R. temporaria*) was used as a random effect. To control for the effect of sample size in the genetic analysis, these models also allowed for the number of individuals analyzed per population. All models were estimated using the function *lme* of the ‘nlme’ package (Pinheiro et al., 2023) in R (R Core Team, 2023).

We also calculated Pearson's correlation coefficient between nucleotide diversity and species richness. However, before calculating the correlations, we first regressed nucleotide diversity and species richness on sample size for the genetic analysis and total abundance in the ponds, respectively. The residuals of these models (i.e., the residuals of the regression analysis between nucleotide diversity and sample size and the residuals of the regression analysis between species richness and total abundance) were then used to estimate the correlation for each focal species (using the *cor.test* function in R). This procedure was used to consider the possible effects of the above-mentioned regressors on nucleotide diversity and species richness.

In all the species-genetic diversity correlations estimated above, we also used evenness (Pielou) and Shannon diversity index as our estimate of diversity. However, none of these correlations were significant (results not shown).

Species richness: correlation with explanatory variables and abundance

We also explored the relationship between species richness and an extended set of environmental, spatial, and land use variables. We tested the three sets of explanatory variables as predictors of species richness *per se*: environmental, spatial, and land use. The environmental set included the following pond characteristics: pond depth, pH, total organic carbon, total nitrogen, total phosphorus, percentage of floating vegetation, percentage of emerging vegetation, and the percentage of bushes surrounding the ponds. In this set, we also included the logarithm of the total abundance of invertebrates as a predictor variable. Spatial variables were generated using the geographic coordinates of the ponds and distance-based Moran's Eigenvector Maps (Dray et al., 2006). Land use variables were the proportions of artificial surface, cultivated area, tree cover, shrub, marsh-peatland, natural surface, and water around ponds (considering a circular buffer of 250 m). We used a forward selection procedure (Blanchet et al., 2008) to select regressors, within each set of explanatory variables, for species richness analyses. The forward-selected variables were used in linear regression models and variation partitioning to estimate their relative contributions to variation in species richness. In these analyses, the importance of forward-selected variables in accounting for species richness was compared to that of the log-transformed pond area, given that ecosystem area is considered a significant determinant of species richness (Drakare et al., 2006; Goorah and Chase, 2020). Pond area was log-transformed to meet the assumptions of normality.

To generate Moran's Eigenvectors maps, we used the *pcnm* function of the R package 'vegan' (Oksanen et al., 2022). To perform forward selection of the explanatory variables, we used the *forward.sel* function of the R package 'adespatial' (Dray et al., 2023). For multiple regression models and variation partitioning, we used the functions *lm* and *varpart* from the 'stats' (R Core Team, 2023) and 'vegan' packages, respectively.

Results

Species-genetic diversity correlations

Despite the large variation in species richness (ranging from 1 to 26 taxa, see Appendix A: Table T1), this variable was not related to genetic diversity (π ; Table 1, Fig. 2). Similarly, π was not related to the different

Table 1

Results of linear mixed effects modeling showing the relationship between nucleotide diversity (π) and different explanatory variables (species richness, pond area (m^2), floating vegetation (%), artificial surface around ponds (%), and total abundance of invertebrates). The number of individuals included in the genetic analysis was also included in the models. Shown here are the estimated fixed effects, standard errors (SE), Student's *t* statistics, and the *P*-values.

Explanatory variable	Coefficient	SE	<i>t</i>	<i>P</i>
(Intercept)	0.1300	0.0428	3.038	0.0045
Species richness	0.0006	0.0015	0.382	0.7047
Number of individuals	0.0016	0.0039	0.408	0.6859
(Intercept)	0.0915	0.0515	1.775	0.0846
Pond area	0.0064	0.0061	1.041	0.3051
Number of individuals	0.0026	0.0037	0.708	0.4836
(Intercept)	0.1410	0.0276	5.110	0.0000
Floating vegetation	-0.0006	0.0024	-0.246	0.8070
Number of individuals	0.0018	0.0029	0.617	0.5411
(Intercept)	0.0911	0.0328	2.778	0.0087
Total abundance	0.0109	0.0071	1.538	0.1330
Number of individuals	0.0020	0.0028	0.713	0.4805
(Intercept)	0.1367	0.0274	4.994	0.0000
Artificial surface	0.0001	0.0006	0.123	0.9028
Number of individuals	0.0019	0.0032	0.594	0.5566

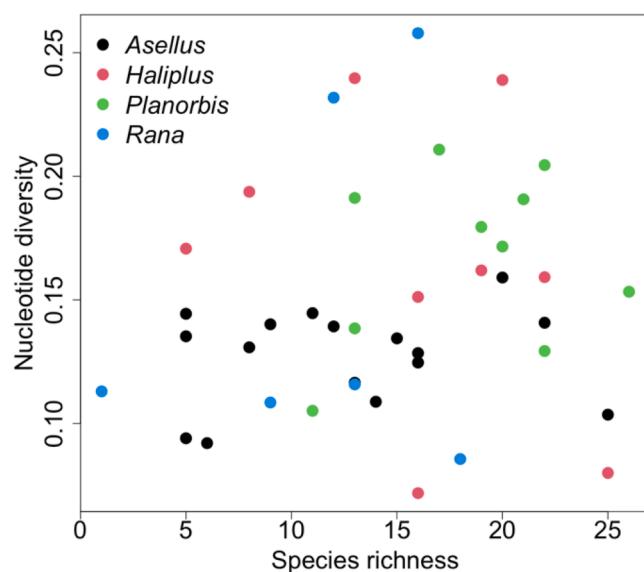


Fig. 2. Relationship between species richness and genetic diversity for different focal species (see text for statistical analyses).

explanatory variables we tested (Table 1, Fig. 3). Consequently, Pearson's correlation coefficient between π of each focal species and the species richness was also not significant (Fig. 4).

Species richness: correlation with explanatory variables and abundance

None of the land use and spatial variables (i.e., Moran's Eigenvector Maps) were selected as potential predictors of species richness according to the forward selection procedure in the linear regression models. Among the environmental variables, only the total abundance of invertebrates was selected by this procedure. The multiple regression model incorporating both total abundance and pond area was statistically significant ($F_{2,28} = 18.06; P < 0.001$) and explained 53.21% of the variation in species richness. Both the total abundance ($t = 3.21; P = 0.003$) and pond area ($t = 2.83; P = 0.008$) were, as expected, positively and significantly correlated with species richness (partial regression coefficients: $b = 2.57 \pm 0.80$ SE and $b = 1.91 \pm 0.68$ SE, respectively; Fig. 5). The residuals of this model were normally distributed (Shapiro-Wilk's $W = 0.90; P = 0.524$) and spatially independent (as indicated by a spatial correlogram where all Moran's I coefficients were not significant; smallest *P*-value = 0.176). This model was also not affected by multicollinearity issues according to the variation inflation factor (VIF) analysis (both variables with $VIF < 1.5$). The unique fractions (adjusted coefficients of determination, R^2) attributed to total abundance and pond area were, respectively, equal to 15.0% and 11.3%, respectively. Interestingly, the largest fraction (adjusted coefficient of determination = 26.9%) was attributable to the shared component.

Discussion

We found no significant relationship between species diversity and genetic diversity in our urban pond system. This result corresponds to the review by Lamy et al. (2017) who found that only 11 and 6 percent of their reviewed studies showed significant positive or negative SGDCs respectively. In our study there could be several reasons for this lack of correlation.

Environmental variables

First, if environmental variables affect SD and GD differently, this could cause a negative or a non-significant SDGD relationship. The

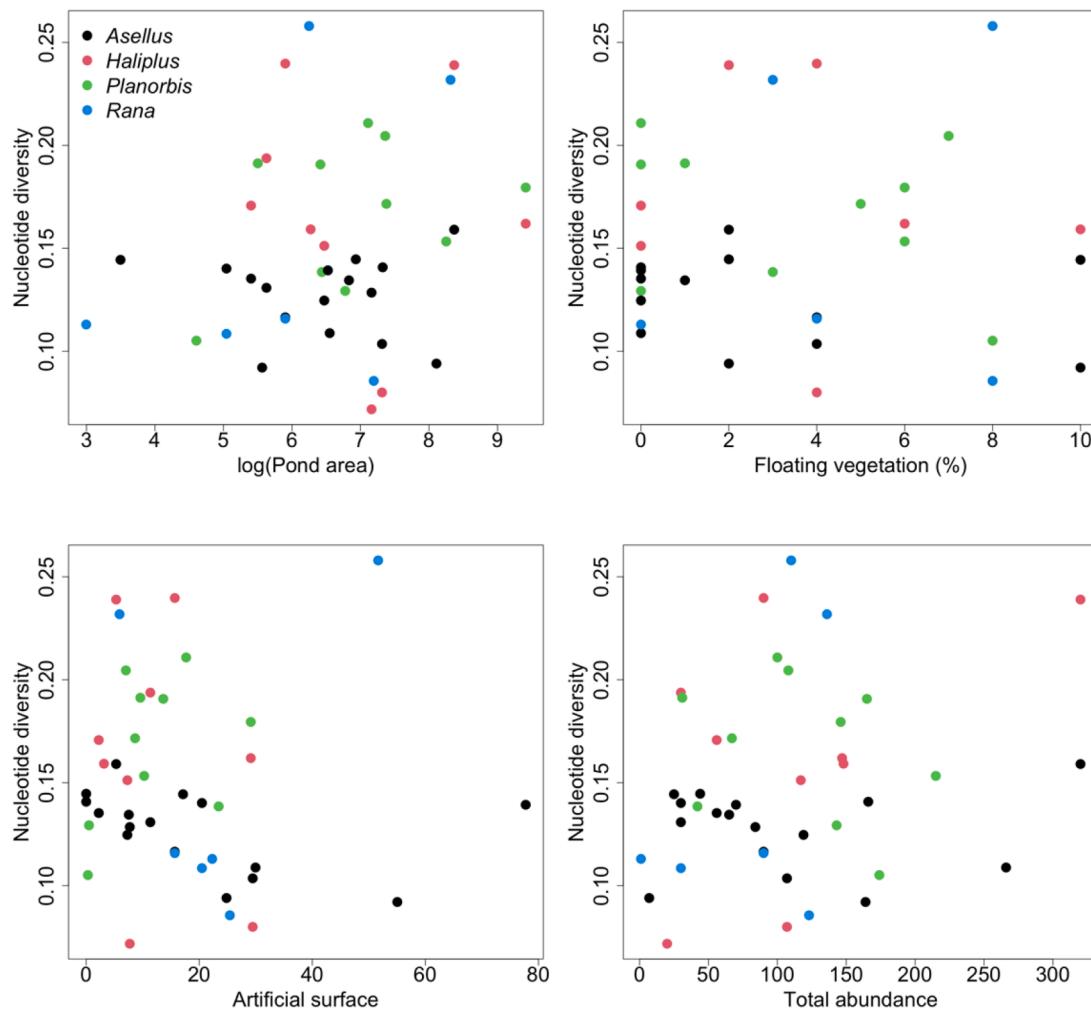


Fig. 3. Relationships between genetic diversity and different explanatory variables (pond area, floating vegetation cover, artificial area surrounding ponds, and total abundance) for different focal species (see text and Table 1 for statistical analyses).

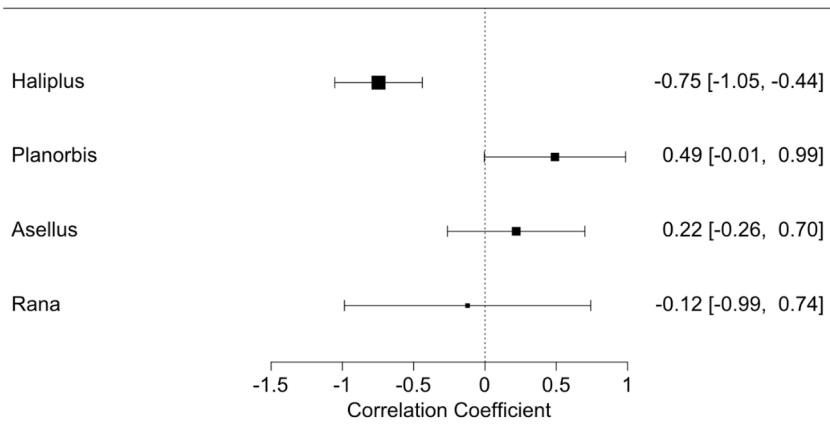


Fig. 4. Pearson's correlation coefficients between species richness and genetic diversity (π) for each focal genus studied (after accounting for total abundance in the ponds and sample size for the genetic analysis, respectively). The size of the symbol is proportional to the sample size. The correlation coefficients are shown on the right side of the plot, as well as their 95% confidence intervals (brackets).

variables we explored — pond area, floating vegetation in the pond, total invertebrate abundance, and artificial surface area around the pond — were not related to GD. On the contrary, these variables have been shown to be significantly correlated with SD, estimated as species richness, in several studies of pond invertebrates in urban ponds (Blicharska et al., 2016; Heino et al., 2017; Hill et al., 2015).

Invertebrate abundance has been shown to be positively correlated with species richness in ponds of the metropolitan area of Stockholm (Heino et al., 2017; see also below). Thus, we expected a positive SGDC given that the focal species also have high abundances. The reason is that, in general, a higher abundance of a species at a site should be correlated with a higher GD (Kimura & Crow, 1964). Although we did

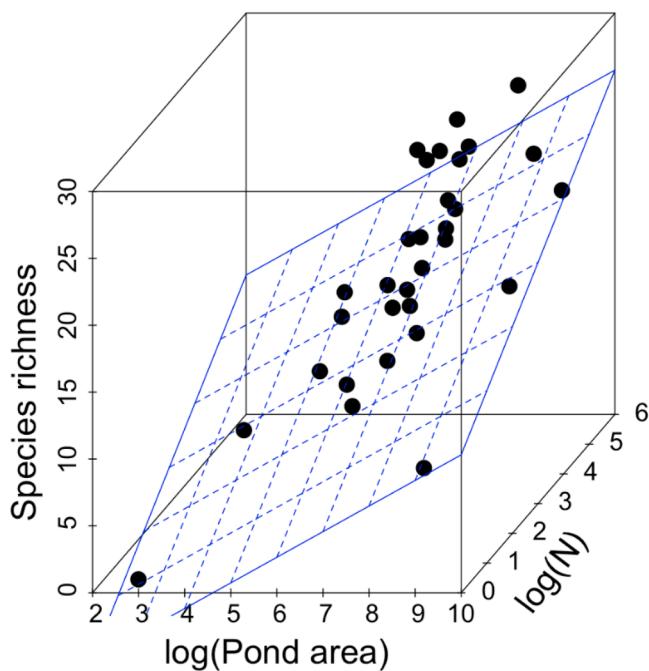


Fig. 5. Relationship between species richness, pond area, and total abundance of invertebrates (N). Both the pond area and the total abundance were log-transformed. The fitted plane is shown in blue.

not find positive SGDC, it is possible that our focal species did not follow the trend of being more abundant in ponds with a high SD. Unfortunately, we do not have enough data on abundance of the four species that we used to calculate GD. An alternative but not mutually exclusive explanation could be that the focal species, if they are somewhat specialized, might have different niche requirements compared to the other species in the community. Such explanations have been suggested by other studies (Ishii et al., 2022; Lamy et al., 2017; Reisch & Scmid, 2019). We note, however, that *Asellus aquaticus* is categorized as a generalist with regard to environmental conditions and resource use (Moog, 2002).

Dispersal effects

SGDC is affected not only by the abundance of the focal species used for GD estimates. It can also be affected by the dispersal properties of the focal species used for GD estimates (Ishii et al., 2022). We predicted that species with intermediate dispersal abilities should show a positive SGDC and that species with a low dispersal ability should show a negative SGDC. This prediction assumes no competition for resources. However, we did not find any support for this prediction.

There could be several reasons for this lack of support. It could be that competition is important in our pond system. Unfortunately, the absence of abundance estimates prevents us from providing mechanistic explanations based on competition effects. Our results may also be due to a small sample size. For example, for the low-dispersal species, we had few individuals for the GD estimates and also few ponds (Appendix A: Table T1). Finally, it could also be that our categorization of dispersal was wrong. In summary, future studies should put more effort into quantifying abundances of the focal species, making it possible to focus more on niche features and competitive effects. An alternative approach would be to focus on functional traits of species to explore the importance of interspecific interactions (Ishii et al., 2022). Such an approach also needs to estimate the association between functional traits and environmental variables in the study system.

Species richness: correlation with explanatory variables and community abundance

Previous studies indicate that local environmental factors and land use variables are significant correlates of species richness in the ponds of the studied region (Blicharska et al., 2016; Heino, et al., 2017). Therefore, it was surprising that few of these variables were significantly correlated with species richness in this study. It should be noted that this study included fewer ponds in the analysis compared to previous studies in the same urban area, which may have affected the results. However, here we also included the total abundance of invertebrates as a predictor of species diversity. According to Gooriah and Chase (2020), to separate the factors underlying species-area relationships (random sampling effects and habitat heterogeneity), one should test the effect of area on species richness, after accounting for abundance. Their study shows that the species-area relationships for lake zooplankton communities in Europe and North America were mainly driven by sampling effects, because, after accounting for abundance, species richness was independent of lake area. Our results are partially in line with those of Gooriah and Chase (2020), as total abundance (after accounting for the pond area) and the shared component (the variation in species richness explained by total abundance, which is in turn related to the pond area) were more important predictors than the pond area (after accounting for total abundance). However, we found that the unique effect of the pond area (i.e., after controlling for variation in total abundance) was still significant and high, explaining 11.3% of the variation in species richness. Therefore, assuming the pond area as a proxy for habitat heterogeneity, our results are also in line with the view that habitat heterogeneity is a consistent predictor of species richness (Ortega et al., 2018; Stein et al., 2014; see also Gooriah et al. 2021 for a meta-analysis that included data on assemblages of plants, invertebrates, herpetofauna, birds and mammals).

Conclusions

Despite recent progress, genetic diversity of populations is seldom given explicit consideration in conservation prioritization (Kahilainen et al., 2014; Xie et al., 2023). This is unfortunate because genetic diversity is important for conservation biology (Kahilainen et al., 2014). If species–genetic diversity correlation (SGDC) is positive, conservation efforts targeting SD or GD would provide similar results on the two. However, as we found here and as has been suggested by several reviews, SGDC is not a general pattern (Lamy et al., 2017). This is not good news for conservation biology because the resources used to benefit SD diversity do not necessarily result in beneficial results for GD. Hence, extra resources must sometimes be used to benefit both estimates of biodiversity. However, more studies are needed, because without replicate studies we will not know whether our results represent a general pattern for urban pond environments.

Data accessibility

The demultiplexed data is archived in the NCBI Sequence Read Archive BioProject ID: PRJNA1026545PRJN, link: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA1026545?reviewer=k77s47nm26he873ahrn69s8jslhttp>

CRediT authorship contribution statement

Frank Johansson: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Yeserin Yildirim:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Chaz Hyseni:** Writing – review & editing, Visualization,

Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jani Heino:** Writing – review & editing, Supervision, Conceptualization. **Jacob Höglund:** Writing – review & editing, Supervision, Conceptualization. **Luis Mauricio Bini:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by grant 2018–00538 from FORMAS to FJ. We acknowledge the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725, for awarding this project access to the LUMI supercomputer, owned by the EuroHPC Joint Undertaking and hosted by CSC (Finland) and the LUMI consortium. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala (# UC-2892). The facility is part of the National Genomics Laboratory (NGI) of Sweden and the Science for Life Laboratory. The SNP&SEQ platform is also supported by the Swedish Research Council and Knut and the Alice Wallenberg Foundation. Many thanks to Ruth Hobro, Jan Pröjts, Jonas Roth, and Hans Erik Wanntorp, for help with species determination. We also want to thank Björn Almqvist, Arvid de Jong, August Lundholm, and Roberta Hedberg for help with field data collection. We also thank Patrik Rödin Mörch, Loïs Rancilhac, Marcel Martin, Pontus Larsson, Lars Littmann and SciLifeLab for advice on data analysis.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.baae.2024.07.002](https://doi.org/10.1016/j.baae.2024.07.002).

References

- Angel, S., Parent, J., & Civco, D. (2012). The fragmentation of urban landscapes: Global evidence of a key attribute of the spatial structure of cities, 1990–2000. *Environ. Urban.*, *24*, 249–283.
- Biggs, J., Williams, P., Whitfield, M., Nicolet, P., & Weatherby, A. (2005). 15 years of pond assessment in Britain: Results and lessons learned from the work of pondconservation. *Aquat. Conserv. Mar. Freshw. Ecosyst.*, *15*, 693–714.
- Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Forward selection of explanatory variables. *Ecology*, *89*, 2623–2632.
- Blicharska, M., Andersson, J., Bergsten, J., Bjalke, U., Hilding-Rydevik, T., & Johansson, F. (2016). Effects of management intensity, function and vegetation on the biodiversity in urban pond. *Urban. For. Urban. Green.*, *20*, 103–112.
- Boda, P., Horváth, G., Kriska, G., et al. (2014). Phototaxis and polarotaxis hand in hand: Night dispersal flight of aquatic insects distracted synergistically by light intensity and reflection polarization. *Naturwissenschaften*, *101*, 385–395. <https://doi.org/10.1007/s00114-014-1166-2>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.
- Bowen, B. W. (1999). Preserving genes, species, or ecosystems? Healing the fractured foundations of conservation policy. *Mol. Ecol.*, *8*, S5–S10.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Mol. Ecol.*, *22*, 3124–3140.
- Di Marsico, M., Cerruti, E., Comino, C., Porceddu, A., Acquadro, A., Capomaccio, S., Marconi, G., & Albertini, E. (2020). MCSeEd (Methylation Context Sensitive Enzyme ddRAD): A New Method to Analyze DNA Methylation. *Methods Molecular Biol.*, *2093*, 47–64.
- Drakare, S., Lennon, J. J., & Hillebrand, H. (2006). The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecol. Lett.*, *9*, 215–227. <https://doi.org/10.1111/j.1461-0248.2005.00848.x>
- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Modell.*, *196*(3–4), 483–493.
- Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guénard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., & Wagner, H. H. (2023). adespatial: Multivariate Multiscale Spatial Analysis. *R Package Version 0.3-21*. <https://CRAN.R-project.org/package=adespatial>.
- Dudgeon, D. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenge. *Biol. Rev.*, *81*, 163–182.
- Duncan, S.I., Robe, E., Rtsø, P. N., Fletche, R. R., Jr, J., & Au, J stin, D. (2020). Urbanization and Population Genetic Structure of the Panama City crayfish (*Procambarus econfinae*). *J. Heredity*111:204–215.
- Excoffier, L., Laval, G., & Schader, S. (2005). Arlequin ver. 3.0: An Integrated Software Package for Population Genetics Data Analysis. *Evol. Bioinform. Online*, *1*, 47–50. <https://doi.org/10.1177/11769343050100003>
- Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annu Rev. Ecol. Evol. Syst.*, *34*(1), 487–515.
- Gál, B., et al. (2019). The effect of urbanization on freshwater macroinvertebrates – Knowledge gaps and future research directions. *Ecol. Indic.*, *104*, 357–364.
- Goertzen, D., & Suhli, N. F. G. (2013). Promoting dragonfly diversity in cities: Major determinants and implications for urban pond design. *J. Insect Conserv.*17: 399–409.
- Gooriah, L.D., & Chase, J.M. (2020). Sampling effects drive the species-area relationship in lake zooplankton. *Oikos*129: 124–132.
- Gooriah, L., Blows, S.A., Sagois, A., Schrader, J., Karger, D.N., Kreft, H., & Chase, J.M. (2021). Synthesis reveals that island species-area relationships emerge from processes beyond passive sampling. *Global Ecol. Biogeography*, *30*(10), 2119–2131.
- Heino, J., et al. (2017). Unravelling the correlates of species richness and ecological uniqueness in a metacommunity of urban pond insects. *Ecol. Indic.*, *73*, 422–431.
- Hill, M. J., Mathers, K. L., & Wood, P. J. (2015). The aquatic macroinvertebrate biodiversity of urban ponds in a medium-sized European town (Loughborough, UK). *Hydrobiologia*, *760*, 225–238.
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, New Jersey: Princeton University Press.
- Hyseni, C., Heino, J., Bini, L. M., Bjalke, U., & Johansson, F. (2021). The importance of blue and green landscape connectivity for biodiversity in urban ponds. *Basic Appl. Ecol.*, *57*, 129–145.
- Ishii, N. I., Hirota, S. K., Matsuo, A., Sato, M. P., Sasaki, T., & Suyama, Y. (2022). Species–genetic diversity correlations depend on ecological similarity between multiple moorland plant species. *Oikos*, *2022*, E09023. <https://doi.org/10.1111/oik.09023>
- Kahlainen, A., Puurinen, M., & Kotiaho, J. S. (2014). Conservation implications of species–genetic diversity correlations. *Glob. Ecol. Conserv.*, *2*, 315–323. <https://doi.org/10.1016/j.gecco.2014.10.013>
- Kimura, M., & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, *49*, 725–738.
- Lamy, T., et al. (2013). Variation in habitat connectivity generates positive correlations between species and genetic diversity in a metacommunity. *Mol. Ecol.*, *22*, 4445–4456.
- Lamy, T., Laroche, F., David, P., Massol, F., & Jarne, P. (2017). The contribution of species–genetic diversity correlations to the understanding of community assembly rules. *Oikos*, *126*, 759–771. <https://doi.org/10.1111/oik.03997>
- Liu, Z., Zhou, T., Heino, J., Castro, D. M. P., Cui, Y., Li, Z., Wang, W., Chen, Y., & Xie, Z. (2022). Land conversion induced by urbanization leads to taxonomic and functional homogenization of a river macroinvertebrate metacommunity. *Sci. Total Environ.*, *825*, 153940.
- MacArthur, R. H., & Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton, New Jersey: Princeton University Press.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.*, *17*(1), 10–12.
- Marconi, G., Capomaccio, S., Comino, C., Acquadro, A., Portis, E., Porceddu, A., & Albertini, E. (2019). Methylation content sensitive enzyme ddRAD (MCSeEd): A reference-free, whole genome profiling system to address cytosine/adenine methylation changes. *Sci. Rep.*, *9*(1), 14864.
- R, K. aps, B. a McKinney, R. A., Courri, R., & oyer, M. (2011). Wetlands as habitat in urbanizing landscapes: Patterns of bird abundance and occupancy. *Landsc. Urban. Plan.*, *100*, 144–152.
- Hill, M. J., Ryves, D. B., White, J. C., & Wood, P. J. (2016). Macroinvertebrate diversity in urban and rural ponds: Implications for freshwater biodiversity conservation. *Biol. Conserv.*, *201*, 50–59.
- Moog, O. (2002). [ed.]. Fauna aquatica Austriaca, edition 2002. Wasserswirtschaftskataster, Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft. [A Comprehensive Species Inventory of Austrian aquatic Organisms With Ecological notes, Edition 2002.]
- Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press.
- Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoechs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M., Lahti, L., McGlinn, D., Ouellette, M., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C., & Weedon, J. (2022). vegan: Community Ecology Package. *R Package Version 2.6-4*. <https://CRAN.R-project.org/package=vegan>
- Ortega, J. C., Thomaz, S. M., & Bini, L. M. (2018). Experiments reveal that environmental heterogeneity increases species richness, but they are rarely designed to detect the underlying mechanisms. *Oecologia*, *188*, 11–22.
- Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods Ecol. Evol.*, *8*, 1360–1373.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS. One*, *7*, e37135.

- Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *J. Theor. Biol.*, *13*, 131–144. [https://doi.org/10.1016/0022-5193\(66\)90013-0](https://doi.org/10.1016/0022-5193(66)90013-0)
- Pinheiro, J., Bates, D., & R Core Team. (2023). nlme: Linear and Nonlinear Mixed Effects Models. *R Package Version 3.1-162*. <https://CRAN.R-project.org/package=nlme>.
- QGIS Development Team. (2021). QGIS Geographic Information System. <https://www.qgis.org>.
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.
- Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conserv. Biol.*, *17*, 230–237.
- Reisch, C., & Schmid, C. (2019). Species and genetic diversity are not congruent in fragmented dry grasslands. *Ecol. Evol.*, *9*, 664–671.
- Rochette, N., & Catchen, J. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.*, *12*, 2640–2659.
- Shannon, C. E., & Weaver, W. (1963). *The Mathematical Theory of Communication*. University of Illinois Press.
- Sinsch, U. (1990). Migration and orientation in anuran amphibians. *Ethol. Ecol. Evol.*, *2*, 65–79.
- Stein, A., Gerstner, K., & Kreft, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecol. Lett.*, *17*, 866–880.
- Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*, 437–460.
- Bat, t, L. y, D. R. eath, G., Fri, N. berg, R. Thornhill, I., Le, M., & dger, E. (2017). Local and landscape scale determinants of macroinvertebrate assemblages and their conservation value in ponds across an urban land-use gradient. *Biodivers. Conserv.*, *26*, 1065–1086
- Till, B. J., Jankowicz-Cieslak, J., Huynh, O. A., Beshir, M. M., Laport, R. G., & Hofinger, B. J. (2015). *Low-cost Methods for Molecular Characterization of Mutant plants: Tissue desiccation, DNA Extraction and Mutation discovery: Protocols*. Cham Heidelberg New York Dordrecht London: Springer. https://doi.org/10.1007/978-3-319-16259-1_9
- Vellend, M. (2003). Island biogeography of genes and species. *Am. Nat.*, *162*, 358–365.
- Vellend, M. (2005). Species diversity and genetic diversity: Parallel processes and correlated patterns. *Am. Nat.*, *166*, 199–215.
- Xie, L., Feng, Y., Zhao, R., et al. (2023). Positive relationships between species diversity and genetic diversity on a local scale at Mt. Wu Yi, China. *Biodivers. Conserv.*. <https://doi.org/10.1007/s10531-023-02695-8>
- Zuur, A. F., et al. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer.