Received: 8 October 2022 | Revised: 27 April 2023 | Accepted: 30 April 2023

DOI: 10.1111/gcb.16757

Global Change Biology WILEY

RESEARCH ARTICLE

Integrative population genetics and metagenomics reveals urbanization increases pathogen loads and decreases connectivity in a wild bee

Katherine D. Chau | Farida Samad-zada | Evan P. Kelemen | Sandra M. Rehan

Department of Biology, York University, Toronto, Canada

Correspondence
Sandra M. Rehan, Department of Biology, York University, Toronto, Canada.
Email: sandra.rehan@gmail.com

Funding information
National Science Foundation; NSERC; Weston Family Foundation

Abstract

As urbanization continues to increase, it is expected that two-thirds of the human population will reside in cities by 2050. Urbanization fragments and degrades natural landscapes, threatening wildlife including economically important species such as bees. In this study, we employ whole genome sequencing to characterize the population genetics, metagenome and microbiome, and environmental stressors of a common wild bee, *Ceratina calcarata*. Population genomic analyses revealed the presence of low genetic diversity and elevated levels of inbreeding. Through analyses of isolation by distance, resistance, and environment across urban landscapes, we found that green spaces including shrubs and scrub were the most optimal pathways for bee dispersal, and conservation efforts should focus on preserving these land traits to maintain high connectivity across sites for wild bees. Metagenomic analyses revealed landscape sites exhibiting urban heat island effects, such as high temperatures and development but low precipitation and green space, had the highest taxa alpha diversity across all domains even when isolating for potential pathogens. Notably, the integration of population and metagenomic data showed that reduced connectivity in urban areas is not only correlated with lower relatedness among individuals but is also associated with increased pathogen diversity, exposing vulnerable urban bees to more pathogens. Overall, our combined population and metagenomic approach found significant environmental variation in bee microbiomes and nutritional resources even in the absence of genetic differentiation, as well as enabled the potential early detection of stressors to bee health.

KEYWORDS
microbiome, network analysis, pathogens, *Ceratina calcarata*, population genomics, urban heat island, wild bees

1 | INTRODUCTION

Rapid expansion of human populations continues to reshape landscapes and poses one of the largest threats to natural ecosystems and biodiversity (Elmgvist et al., 2021; Otto, 2018). To compensate, it is expected that urbanization will drive over two-thirds of the human population to reside in cities by 2050 (Ritchie & Roser, 2018). In this study, we focus on different genetic and environmental

aspects influencing organism health, by analysing (1) population genetics, (2) metagenomics, (3) microbiome composition, and (3) the integration of population genetics and metagenomics. We also emphasize the importance of urbanization as it relates to these topics in the context of wild bee health.

Urbanization propagates climate change in various ways such as contributing close to 70% of global carbon emissions (Elmgvist et al., 2021), degrading and fragmenting natural habitats, and introducing non-native species, all of which increase the number of at-risk species and local extinctions (Des Roches et al., 2020; Ouyang et al., 2018). As a result, research on the conservation and ecology of species has recently expanded to include studies of species diversity, adaptation, genetics, and evolution in an urban context (Ouyang et al., 2018). Compared to rural environments, cities have multiple features that impose different environmental pressures relative to rural habitats. For example, the urban heat island effect (i.e., higher temperatures in cities than in their surroundings) interferes with plant phenologies, affecting flowering times and length of growing seasons (Dhami et al., 2011; Kabano et al., 2021). Of particular concern are the impacts cities have on insect pollinators, one of the most important groups of organisms that are responsible for pollinating over 87% of flowering plants and 75% of global food crops (Aizen et al., 2009; Klein et al., 2007; Ollerton et al., 2011). Bees are the most prominent insect pollinators that has led to extensive research and management of bees for agricultural purposes (Aizen et al., 2009; Kleijn et al., 2015). However, only a handful of species are managed, whereas the majority of wild bee species are unmanaged, and many of the wild bees are declining at alarming rates (Bartomeus et al., 2013; Danforth et al., 2019; Mathiasson & Rehan, 2019; Ollerton et al., 2011; Potts et al., 2016). Urbanization and agricultural practises lead to pollution, floral homogenization, pesticide run-off, habitat fragmentation and destruction that contribute to wild bee declines (Cardoso & Gonçalves, 2018; Odanaka & Rehan, 2019; Pereira et al., 2021).

As an effort to maintain bee diversity within urban landscapes, recent studies on urban green spaces found they may act as reservoirs for bee fauna (Banaszak-Cibicka et al., 2018; Hülsmann et al., 2015; Theodorou, Albig, et al., 2016), although this may favour exotic species (Fitch et al., 2019; Geslin et al., 2020; Russo et al., 2021). Additionally, parasitism and pathogen infections in bees are also a major driver in global bee population declines, a problem enhanced by urbanization and habitat degradation. This happens because the interchange of goods, people, and domestic animals in cities can introduce/spread pathogens and parasites (Youngsteadt et al., 2015). For instance, one study found that while urban areas had greater bumble bee abundance due to greater floral diversity, increased urbanization simultaneously heightened abundance of a protist, *Crithidia bombi*, which parasitizes bumble bees (Theodorou, Radzevičiūtė, et al., 2016). Urbanization can also propagate disease spill over, where infected individuals visit flowers that in turn become contaminated with the pathogen or parasite of the infected individual and can spread to other healthy individuals, populations, or even cross between bee species (Proesmans et al., 2021; Tehel et al., 2016).

Studies on bees in urban settings have documented various shifts in their ecology, such as behavioral, morphological, and physiological changes. For example, decreased thermal tolerance was more pronounced in urban bees than rural bees due to the urban heat island effect (Burdine & McCluney, 2019; Sánchez-Echeverría et al., 2019). Cities support larger bees as they can forage farther and longer in a highly fragmented environment (Austin et al., 2022; Cardoso & Gonçalves, 2018; Theodorou et al., 2021); thus, smaller-bodied bee species, many of which are typically solitary species, are susceptible to limited gene flow and reduced reproduction in urban environments where foraging and nesting grounds are further apart (Austin et al., 2022; Zurbuchen et al., 2010). Conversely, urban warming in cities may also favour smaller bee species as they have lower risks of overheating (Theodorou et al., 2021). Still, studies find that overall, bee abundance declines as urban warming increases (Hamblin et al., 2018).

With increases in urbanization, the effect of environmental degradation greatly impacts wild bees, yet studies have not documented the population genomics, microbiomes, or possible disease vectors for most species. The lack of a basic understanding of wild, solitary bee health poses a great risk to global food security and diverse ecosystem health, since close to 85% of all described bee species are solitary (Batra, 1984) and provide the majority of pollination services (Batra, 1984; Breeze et al., 2011; Magnacca & Brown, 2012). Urbanization also impacts the population dynamics of wild bees such that the integration of population genomics is essential in order to fully capture how urban environments are impacting the health and fitness of bees and their populations For example, we can identify critical corridors for gene flow in bee populations found across variable landscapes by integrating information from genetic structure and land use (Jha, 2015; Jha & Kremen, 2013), or detect key genes under selection that may explain the environmental stressors on a bee population (Kent et al., 2018; Theodorou et al., 2018). Furthermore, population genetics can also reveal inbreeding evidence in bees living in cities or whose habitats have urbanized over time (Austin et al., 2022).

Advances in next-generation sequencing have prompted the use of metagenomic data, allowing the genetic material of coexisting taxa to be sequenced in concert with the species of interest (Hugenholtz & Tyson, 2008). Bees are associated with beneficial microfauna (i.e., microbiome which include bacteria, viruses, protists, and fungi, Engel et al., 2016), some of which are essential for bee development (i.e., core taxa) while others can be harmful or pathogenic (Engel et al., 2016; Graystock et al., 2017). Characterization of non-pathogenic gut microbial and fungal communities revealed their important roles in bee nutrition and digestion (Engel et al., 2012; Kwong et al., 2014), immunity (Lang et al., 2022; Raymann & Moran, 2018), development (Dharampal et al., 2019; Kapheim et al., 2015; Nguyen & Rehan, 2022a; Tarpy et al., 2015), and cognitive ability (Li et al., 2021). Similar to microbiomes, bee metagenomes reveal many important environmental relationships and include taxa

not typically studied in microbiome studies (e.g., plants, arachnids, and nematodes). For example, the detection of specific plant DNA can highlight crucial floral associations essential for urban green space planning and at-risk bee management (Lawson et al., 2020; Rocha-Filho et al., 2021), identify plants as hubs for potential pathogens (Durrer & Schmid-Hempel, 1994; Graystock et al., 2020), and indicate vectors for bee viruses (Guo et al., 2021; Schoonvaere et al., 2016).

Substantial work integrating urban ecology, population genomics, and metagenomics has been done on corbiculate bees (i.e., *Apis* and *Bombus* species, Austin et al., 2022; Li et al., 2021; Raymann et al., 2018). These studies have been foundational for identifying core bacterial symbionts, detection of pathogens, identifying dietary breadth, and classifying other disease vectors. Conversely, studies on the core microbiomes and metagenomes of wild, solitary bees remain in their infancy. While social corbiculate bee microbiomes are reduced and derived from members in the colony, the dynamics of wild solitary bee microbiomes are diverse as they are highly variable and dependent on the environment (Cohen et al., 2020; Engel et al., 2016; Kapheim et al., 2021).

Here, we examine solitary bee ecology in urban environments by integrating population genomics and metagenomics for the small carpenter bee, *Ceratina calcarata*. This solitary wild bee species is a prime candidate to study the influence of environment on its microbiome and metagenome since it is distributed widely and across variable landscape features (Kelemen & Rehan, 2021). *Ceratina calcarata* is a solitary, native bee to North America that nests in dead, woody stems found in grassy fields, shrub lands, meadows, and open forests (Hanula et al., 2015; Rehan & Richards, 2010), and is among the top 20 bee species with highest mean contribution to crop pollination, making it a highly economically important species (Kleijn et al., 2015). Recently, studies on *C. calcarata* have characterized its microbiome. For instance, Graystock et al. (2017) identified distinct gut microbiota in *C. calcarata* including prominent taxa such as *Lactobacillus*, *Pseudomonas*, *Gilliamella*, and *Snodgrasella*. Similarly, Nguyen and Rehan (2022a) documented changes in microbiome composition at different developmental stages.

This study uses a whole-genome sequencing approach integrating population genomics and metagenomic data to assess the influence of urban environmental features on wild bee genetics and ecological interactions across four main objectives. First, we estimate the genetic diversity and population structure of *C. calcarata* and assess how land use characteristics affect gene flow within an urban setting. We hypothesize that population structure within the city is driven by landscape features and that areas dominated by vegetation facilitate *C. calcarata* dispersal, while impervious surfaces limit gene flow. Second, we determine how differences in landscape influence the metagenomic and microbial diversity of the bee, and we expect differences in taxa diversity across an urbanization gradient. Third, we characterize the core microbiome of *C. calcarata*, and we predict our urban bee will possess different core microbiota when compared to previous bee studies. Fourth, we identify taxa potentially parasitic or pathogenic to urban bees and other wildlife for future management and conservation practises. We predict that urban sites will harbor more potential pathogens relative to rural sites. Our study provides the first whole genome sequencing, population genomic and metagenomic characterization of a wild, solitary bee in an urban context, and documents the complex relationships between bee population structure, metagenomic interactions and landscape in an urban environment.

2 | MATERIALS AND METHODS

2.1 | Sample species, area, and sequencing

We sampled *C. calcarata* from Toronto, Canada (43.6532°N, 79.3832°W; Figure S1). The sample region was mapped using ArcGIS pro 10.7 (ERSI) and was used to overlay transects of 110 × 75 km with 5 km square grid cells. We selected 66 sites (approximately 1–3 bee samples within each site location) that covered a broad urbanization gradient across the sample region Environmental climate for each site was obtained from the WorldClim dataset (https://www.worldclim.org/) with values obtained for the point of collection, and green space and developmental percentages per site was obtained from the Ontario Land Cover Compilation (version 2.0) dataset (https://geohub.lio.gov.on.ca/documents/7aa998fdf100434da27a41f1c637382c/about) with a 500 m buffer around each sample collection location. In total, we collected 180 *C. calcarata* females across the 66 sample sites from August 2019 to October 2020 (Table S1; Figure S1). Nests were collected and dissected, allowing for one adult female from each nest to be preserved in 95% ethanol; nests were not used further in this study.

DNA was extracted from whole bee specimens using Quick-DNA Miniprep Plus extraction kit (Zymo Research) following a modified protocol from Freitas et al. (2020). Qubit Fluorometer (Thermo Fisher Scientific) assessed DNA quantity. For each bee, DNA concentrations were normalized to 20 ng/μL. Library preparation using the NEB Ultra II library kit, and sequencing on Illumina NovaSeq 6000 with pair-end 150 base pair reads were conducted at Génome Québec. Reads were submitted to NCBI short read archive under BioProject PRJNA725238. Reads were checked for quality, trimmed, and mapped against the *C. calcarata* reference genome (PRJNA791561). Additional information on study species and sample data processing are provided in the supplementary methods.

2.2 | Population genomics

Two methods of variant calling, Genome Analysis Toolkit (GATK version 4.2.2.0, McKenna et al., 2010), and bcftools (version 1.4; Li, 2011) were implemented to assess the population structure of

C. calcarata. After intersecting the two single nucleotide polymorphism (SNP) datasets, performing full filtering based on quality, depth, minimum allele count, and missing data percentage; identifying loci not in Hardy–Weinberg Equilibrium (HWE) and potentially paralogous loci (see supplement for details), we used the resulting SNP dataset for population genetics analysis. At various filtering stages, samples with insufficient coverage <3×, siblings and highly inbred individuals were removed ($N = 18$), and 162 samples were retained in the final dataset.

First, we calculated mean observed heterozygosity (H_o), mean expected heterozygosity (H_e), and mean inbreeding coefficient (F_{IS}) using vcftools (version 0.1.16; Danecek et al., 2011). Additionally, using 20,000 randomly selected SNPs, we calculated F_{IS} using 1000 permutations in Genetix version 4.05.2 (Belkhir et al., 2004) to determine levels of significance. We computed Tajima's *D* (Tajima, 1989) and nucleotide diversity by π (Nei & Li, 1979) in bin intervals of 10 kb using vcftools.

To identify the likely number of genetic clusters within our dataset, we used Bayesian analysis of genetic structure as implemented in ADMIXTURE (version 1.3.0; Alexander et al., 2009), employing an unsupervised model and varying the number of clusters (*K*) from one to ten, with ten iterations per *K* value. We used the *K* value with the lowest cross-validation error estimate as our optimal number of clusters.

To understand how geographic distance, land use, and environmental variables affect dispersal in *C. calcarata*, we calculated pairwise genetic, geographic, resistance, and environmental distance matrices between all pairs of samples within our population. To create a genetic matrix, we used Yang's pairwise relatedness (Yang et al., 2010). This metric is commonly used in landscape genetics studies to estimate inter-individual genetic distance that reflects recent gene flow (Dalapicolla et al., 2021; Shirk et al., 2017).

We created a pairwise geographic matrix by computing the shortest path between all pairs of locations as implemented in distm command from R package *geosphere* (Hijmans et al., 2019). For resistance surfaces, we used the raster layer of 2020 land use and land cover at 10 m resolution (ESRI 2020; Karra et al., 2021). Following Ballare and Jha (2021) and Jaffé et al. (2019), we assigned a resistance value of 0.1 to land cover categories that were hypothesized to facilitate *C. calcarata* movement, and resistance value of 0.9 to land cover categories that were hypothesized to limit dispersal. We generated several hypotheses (Table 1), and created resistance surfaces for each, assigning different resistance values for each land cover type, depending on the hypothesis being tested. For the null hypothesis, we assigned a resistance value of 0.5 across all land use categories, creating a raster where dispersal pathways were independent of land use type. For each hypothesis, we created a resistance distance matrix based on least-cost path distances (Adriaensen et al., 2003) using the Dijkstra's algorithm as calculated by the costDistance (*gdistance*) (van Etten, 2017). Additionally, we used the same resistance rasters to calculate pairwise resistance distances between all samples in our population using circuit theory as implemented in Circuitscape version 5.0.0 (Hall et al., 2021; McRae, 2006). Similarly, we tested for isolation by environment (IBE) for two standardized bioclimatic variables (BIO1: Temperature; BIO12: Annual Precipitation) from WorldClim (https://www.worldclim.org/data/bioclim.html).

Finally, we performed multiple regression on distance matrices (MRM) using 1000 permutations between Yang's pairwise relatedness and land use resistance distances for each hypothesis in Table 1, as well as between Yang's pairwise relatedness and environmental distance (IBE), as implemented in R package *ecodist* (Goslee & Urban, 2007). Detailed information for each method is available in the supplement.

2.3 | Metagenomics

We assigned taxonomy to unmapped reads using metaSPAdes (version 3.10.1; Nurk et al., 2017). Resulting contigs for 180 samples were BLASTed using blastn (version 2.12.0; Altschul et al., 1990)

TABLE 1 Results of the isolation-by-resistance (IBR) and isolation-by-environment (IBE) hypothesis testing using least-cost path modelling and circuit theory (Circuitscape) approaches. Circuitscape results are in parentheses.

		Resistance values[b]					*F*	R^2 (%)	*p*-value
Hypothesis #	Model[a]	Urban area	Crops	Grasses, shrubs	Trees	Water, flooded, or bare ground			
0	Null	0.5	0.5	0.5	0.5	0.5	121.26 (158.91)	1.83 (2.38)	0.001 (0.001)
1	**IBR 1**	**0.9**	**0.1**	**0.1**	**0.1**	**0.9**	**155.20** (158.88)	**2.33** (2.38)	**0.001** (0.001)
2	IBR 2	0.9	0.9	0.1	0.1	0.9	151.16 (148.33)	2.27 (2.22)	0.001 (0.001)
3	IBR 3	0.1	0.9	0.1	0.1	0.9	117.58 (117.41)	1.77 (1.77)	0.001 (0.001)
4	**IBR 4**	**0.9**	**0.1**	**0.1**	**0.9**	**0.9**	153.23 **(162.25)**	2.30 **(2.43)**	0.001 **(0.001)**
5	IBR 5	0.9	0.9	0.1	0.9	0.9	129.34 (140.69)	1.95 (2.11)	0.001 (0.001)
6	IBE	NA	NA	NA	NA	NA	138.09	2.07	0.001

Note: Bolded hypotheses had highest R^2 value based on either least-cost path or circuit theory.

[a]IBR model test for effect of land use type. IBE tests for effect of bioclimatic variables (annual temperature and annual precipitation).

[b]Resistance values: 0.1 = low resistance, 0.5 = neutral resistance, 0.9 = high resistance.

against the non-redundant nucleotide *nt* database. We removed contigs shorter than 300 bp and any hits to *Sodalis* or *Wolbachia* intercellular bacteria, as these are considered environmental contaminants (Graystock et al., 2017) (Table S2). For our environmental variables, we examined: (1) annual precipitation, (2) annual temperature, (3) urban developmental percent, and (4) green space percent (Table S3). We chose these four environmental variables as they relate to urbanization. Urban areas typically consist of high development due to the transformation of open green spaces to impervious land cover, and cities harbor distinct microclimates that alter temperature and rainfall within cities rural areas (Ayers & Rehan, 2021; Yang et al., 2021). Each environmental variable was classified into five categories 1—very low, 2—low, 3—moderate, 4—high, and 5—very high (Figure 2; Table S3). Next, we extracted all contigs that matched to any of the seven domains (arachnids, bacteria, fungi, plants, nematodes, protists, and viruses) and filtered the remaining contigs using a minimum relative contig abundance threshold of 0.1% within a sample per domain (Tables S4–S8). Next, we summed contigs from samples belonging to the same site, producing contig data for 66 sites from 180 bee samples.

We tested for effects of spatial autocorrelation with metagenomic composition using family or genus level alpha diversity, compared against pairwise distances between nearest neighbor sites. We implemented non-parametric Spearman rank correlation tests using the COR.TEST method in R with method set to "spearman". Next, we ran multiple linear regression using the LM function, as well as a robust multiple linear regression model using the RLM function from the *MASS* package (Venables & Ripley, 2002) and the F.ROBFTEST from the *sfsmisc* package (Maechler, 2022) to obtain F-statistic and p values, in order to identify relationships between metagenomic alpha diversity with each of the environmental variables using continuous values. We also implemented redundancy analysis (RDA) to model the effects of the environmental variables on metagenomic diversity, using the RDA function in R.

Next, we ran diversity analyses on our combined and individual datasets (i.e., for each domain separately) to assess microbiome and metagenome diversity across the different environmental variable categories. We used R package *vegan* (Oksanen et al., 2020) to run Bray–Curtis dissimilarity and ran PERMANOVA to assess if dissimilarity values had a significant effect on community composition. We then performed multidimensional scaling (MDS) using CMDSCALE method in R, on all environmental variables for the entire dataset and for each domain separately to visualize clusters. All R analyses were done using R Studio, R version 4.2.0 (R Core Team, 2022; RStudio Team, 2021). Next, similarity percentage (SIMPER) in PAST (version 4.06; Hammer et al., 2001) was implemented to identify top 10 taxa contributing most to diversity across sites for each environmental variable category. Finally, we implemented random forest regression and classification analyses using the R package *randomForest* (Liaw & Wiener, 2002) to assess how well microbial community structure can identify patterns in environmental variables. If random forest classification produces very high out-of-bag error rates (Mitchell, 2011) or regression produces negative R^2 variance values (Chicco et al., 2021), this indicates the random forest models were not good fits for predicting environmental feature using metagenomic data and instead, community composition is randomly distributed.

To identify which taxa cluster together depending on environmental variable, we ran negative binomial distribution analysis (NBDA) from the R package *DESeq2* (Love et al., 2014) that was conducted using family level and genus level data from all domains combined (i.e., 'all') and compared between each environmental variable category, resulting in 10 comparisons per environmental variable. Normalized contigs were implemented in weighted gene co-expression network analysis in R (WGCNA; Langfelder & Horvath, 2008), and network analysis were conducted following the methods outlined by Langfelder and Horvath at https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/. We then performed functional analysis using FragGeneScan version 1.31 (Rho et al., 2010) to find fragmented genes in our contigs, and implemented eggNOG-mapper (version 2.1.6; Cantalapiedra et al., 2021) and GhostKOALA (Kanehisa et al., 2016) to identify gene function based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) identifiers and reconstructed pathways. The number of reconstructed pathways is interpreted as productivity. Additional methods on taxonomic classification, network analysis, functional analysis, and diversity analyses are found in the supplement.

2.4 | Microbiome composition

To determine if *Ceratina calcarata* harbors important, known microbiota essential for healthy development, we scanned the microbiome for notable bacteria, viruses, or fungi based on previous studies focused on bee microbiomes. Since bees can act as vectors of plant disease and plants are an essential food source for bees, we also scanned the microbiome for notable plant bacteria, viruses, and fungi based on previous plant studies focused on plant growth and health. Next, family or genera were characterized as core if they were found in at least 50% of the sites (≥33 sites) and present with a relative contig abundance ≥1% within their domain across the entire dataset. We then compared our core and prominent bacteria (i.e., not classified as core in our dataset but are known to be important in bee health) against known core bacteria in other corbiculate and solitary bees, to assess if *Ceratina calcarata* sampled across an urban landscape harbors a shared or unique core bacterial microbiome.

2.5 | Population genomics, landscape and metagenomic correlation analyses

To infer potential risks to wild bee health, we analysed the correlation between population- and landscape-level data, and potential bee and plant pathogen diversity. To obtain pathogen information for each sample comparison, we converted raw pathogen contig

abundances to Bray-Curtis dissimilarities to obtain potential pathogen and parasite diversity (hereafter referred to as pathogenicity). We used CoNet (R package CoNetinR version 0.0.0.9000; Faust & Raes, 2016) to test for correlations between pathogenicity, Yang's relatedness, environmental distance, and resistance distance across 162 pairwise sample comparisons ($N = 13{,}041$). To avoid erroneous network edge scores produced by biases from different correlation and dissimilarity methods, CoNet networks were created using Pearson, Bray-Curtis, Spearman, and Kullback–Leibler edge scores with 100 iterations and multiple-test corrections using Benjamini-Hochberg. Correlations are considered significant with $p < .05$ for comparisons above and below the diagonal for the correlation analysis.

3 | RESULTS

3.1 | Urbanization across the sample region

Our sample region covered a broad range in development, green space, temperature, and precipitation (Figure 1). Within the core of our sample area towards the city center (towards Lake Erie shore), developmental percentages fall within the 80%–100% as human population density is highest in this area (4457 people/km^2) (Figure 1a). Likewise, green space percentages which encompass open green spaces, parks, agricultural land, and forests are between 0%–24% (from very low to moderate levels) (Figure 1b). Outside of the city there are very high levels of green space and very low developmental area. Category 5 of green space percentages ranges from 49% to 86%, thus, majority of sites fall into categories 1–4 which consist of green space percentages falling between 0%–42%. The climatic variables portray a longitudinal and latitudinal pattern. Annual precipitation ranges from 791 mm to 913 mm, with the highest levels of annual precipitation (categories 3–5) occurring to the far east of the sample area (Figure 1c). Towards areas of high development, annual precipitation levels are lower. Finally, temperature follows a distinct latitudinal shift, shifting from very low annual temperatures (60–66.2°C) outside of the sample area, to very high temperatures (81.7–85.3°C) towards highly urbanized areas (Figure 1d).

3.2 | Population genomics of *C. calcarata* and the importance of green spaces for wild bee connectivity

After filtering, we obtained 942,834 polymorphic SNPs across 162 samples, with a mean coverage depth of 9.62× (range 4.4× to 14.8×) and mean missing percentage of 4.20%. Within this dataset, 89.75% of SNPs were located on 16 major scaffolds (an average of 52, 886.9

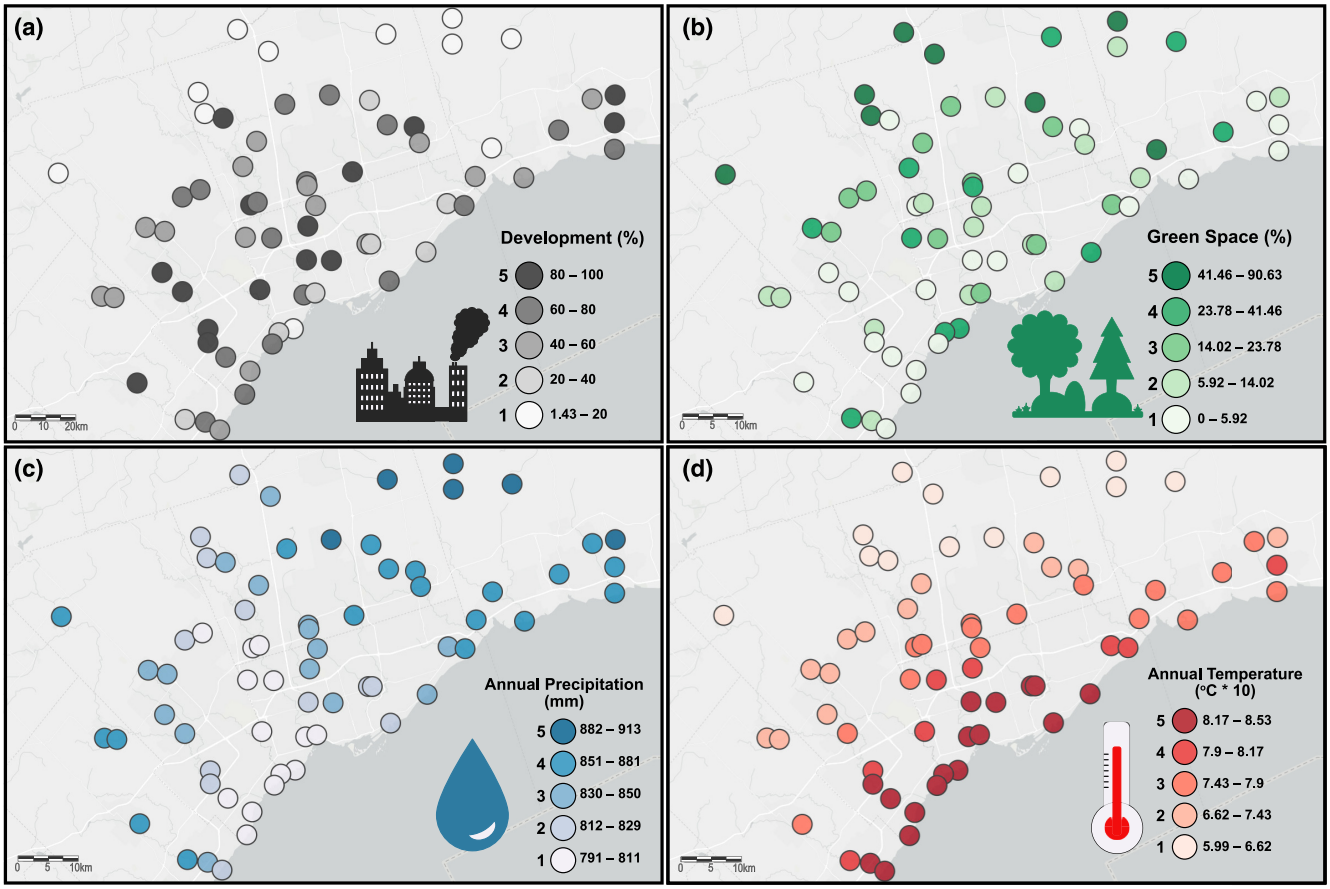

FIGURE 1 Four landscape environmental variables across samples across the urban landscape: (a) developmental percent, (b) green space percent, (c) annual precipitation, and (d) annual temperature.

SNPs per scaffold). Overall, 126,596 SNPs were out of HWE and/or identified as potential paralogs and hence were removed, resulting in 816,238 SNPs that were used for downstream population genetics analyses.

Nucleotide diversity (π) estimate was 0.00075 and the mean F_{IS} was 0.08 and significantly different from 0 ($p < .01$), indicative of high inbreeding ($n = 162$, $H_o = 0.19$, $H_e = 0.21$). Furthermore, we found that mean Tajima's D demonstrated a positive bias across the genome for the 16 major scaffolds (Figure S2), consistent with either a recent population contraction or balancing selection (Tajima, 1989). The results of the ADMIXTURE analysis revealed that all samples in our dataset most likely belong to the same genetic cluster, as $K = 1$ had the lowest cross-validation error (Figure S3).

We found a weak but significant ($R^2 = 0.0183$, $p = .001$, Figure S4) decrease in pairwise relatedness with increased geographic distance, even when all land use categories were given equal resistance values (null hypothesis, Table 1). Across all categories, pairwise relatedness decreased with increased resistance distance, but correlation was higher in scenarios where vegetation rich areas were predicted to facilitate dispersal and urban/impervious surfaces were predicted to limit dispersal, based on values derived by both modelling approaches. While least-cost path modelling found that

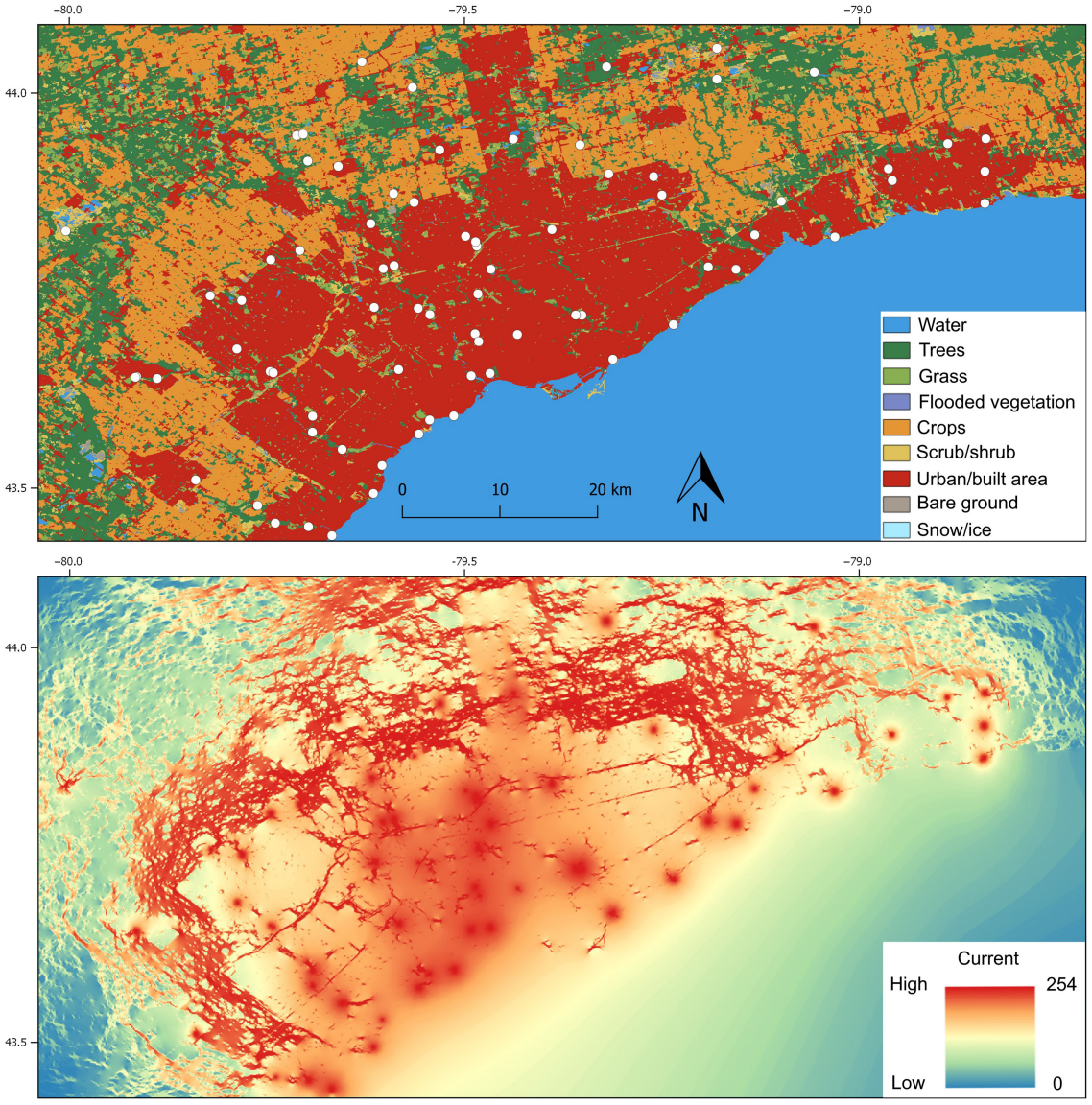

FIGURE 2 Land use across the study area with white circles indicating collection sites (top). Data source: Esri 2020 Land Cover. Karra et al. (2021). Global land use/land cover with Sentinel-2 and deep learning, IEEE, 2021. The following land use classes were present within the examined area: water (32.03%), trees (14.86%), grass (3.44%), flooded vegetation (0.05%), crops (20.38%), scrub/shrub (1.84%), urban/built area (27.11%), and bare ground/ no vegetation (0.29%). Map generated in Circuitscape using resistance values from hypothesis 4 (IBR 4) (bottom). Warmer colors in the bottom map show higher cumulative current flow and correspond to green spaces, including, grasses, crops, and shrub/scrub land based on the top map. Maps produced using QGIS Geographic Information System, Open Source Geospatial Foundation Project (http://qgis.org).

pairwise relatedness is more correlated with resistance distance in a scenario where grasses, shrub/scrub, trees, and crops were given low resistance values (hypothesis 1; Table 1, Figure 2; Figure S5a), Circuitscape results indicated that giving the lowest resistance to only crops, grasses, and shrub/scrub results in the highest correlation with pairwise relatedness (hypothesis 4; Table 1, Figure 2; Figure S5b). Both modelling approaches showed that assigning low resistance values to urban areas (Hypothesis 3) resulted in the lowest correlation with pairwise relatedness across all examined scenarios, including the null hypothesis. Furthermore, Hypothesis 3 was the only scenario where the individual contribution of resistance distance was not significant (Table S9).

3.3 | Taxa family and genus richness driven by densely urban sites

To understand how urbanization impacts bee metagenomic diversity, we analysed the metagenomic composition from all 180 wild bee samples across environmental gradients encompassing 66 sites and the distribution of families and genera belonging to the seven major domains (Figure 1). We tested for effects of spatial autocorrelation on metagenomic composition using nearest neighbor pairwise distances and alpha diversity between the pairwise sites. Alpha diversity for nearest neighbor pairwise distances were not significant at the family ($S=38{,}665$, rho $=0.193$, $p=.1207$) or genus level ($S=40{,}418$, rho $=0.156$, $p=.2101$) (Figure S6). Multiple linear regression models testing metagenomic alpha diversity against each of the environmental variables were not significant (Table S10). Furthermore, our RDA models were not significant at the family or genus level (Table S10).

Both families and genera level alpha diversity, represented by family/genera richness, were highest towards the lake edge, where development and temperature levels are the highest, precipitation is low to moderate, and where green space is the lowest (Figure 3; Figure S6). Family and genera alpha diversity is highest overall and forms a distinct cluster with very high temperatures and development, very low green space, and low to moderate precipitation (Figure 4). Although sites can overlap in environmental features, the number of sites for each environmental feature is different, and we observe that the urban cluster includes features such as high to very high temperature, very low green space, moderate to low precipitation, and high to very high development (Figure 4). In contrast, lowest family and genera alpha diversity clustered at sites where precipitation and green space levels were the highest, and where temperature and development was the lowest.

Overall metagenomic beta diversity was not predicted by any of the four environmental variables (annual precipitation, annual temperature, developmental percent, and green space percent) (PERMANOVA, $p>.05$; Table S13) (Figures S8–S11). Since the random forest classification error rates exceeded 75% and random forest regression R^2 variance values were negative for all environmental variables, this suggests that based on random forest analyses community composition is randomly distributed with (Table S14). However, when examining focal domains, fungal genera and nematode families varied significantly by green space percent coverage (PERMANOVA, fungi $R^2=0.918$, $p=.035$, nematode $R^2=0.709$, $p=.024$). The greatest number of fungal genera and nematode families were found in sites where temperatures and development were high and green space was low (Figure 3).

Negative binomial distribution analyses (NBDA) found significant overrepresentation of the bacterial family Enterobacteriaceae in *C. calcarata* sampled in sites with low annual precipitation compared to sites with high annual precipitation (Table S15). At the highest annual temperatures, four plant families (mustards (Brassicaceae), asters (Asteraceae), mallows (Malvaceae), and grasses (Poaceae)) and the protist family Plasmodiidae were significantly overrepresented (Table S16). Changes in developmental percent did not correlate with any significant overrepresentation of any taxa (Table S17). At low green space cover, legumes (Fabaceae) were significantly overrepresented (Table S18). NBDA found no significant results at the genus level. Weighted gene co-expression network analysis (WGCNA) further corroborated NBDA findings, identifying several key co-occurring plant and bacterial families depending on environmental feature. Plant families Asteraceae, Brassicaceae, and Malvaceae co-occurred with other plants and bacteria when temperatures were high (Figure S12). The bacterial family Enterobacteriaceae had several genera co-occurring with plants when annual precipitation levels are low (Figure S12). Overall, these clusters with significantly co-occurring hub taxa demonstrate that the majority of hub families/genera are bacteria and plants. Based on hierarchal clustering, sites with low annual precipitation levels had several hub bacterial genera form a cluster module that belonged predominantly to Enterobacteriaceae: these included genera such as *Enterobacter*, *Pluralibacter*, and *Klebsiella* (module correlation $=0.37$, $p=.017$, Table S19). Furthermore, Enterobacteriaceae genera co-occurred with plants belonging mainly to Asteraceae at these sample sites. Sites with high annual temperatures identified several hub plant genera clustering together, with the most prominent plant family being Asteraceae, which was identified as an overrepresented plant family from NBDA results (module correlation $=0.69$, $p<.01$, Table S20). Finally, sites with low green space cover identified overrepresentation by legumes (Fabaceae) (Table S18). However, according to WGCNA the majority of plant genera in these sites belonged to mints (Lamiaceae)—there was one overrepresentation of Fabaceae at the genus level (*Acacia*) (module correlation $=0.5$, $p<.01$, Table S22).

When comparing across all domains, SIMPER analyses revealed that among plant families, Asteraceae and Fabaceae had the greatest dissimilarities in relative abundance across categories for each environmental variable (Table S23). Similar to previous NBDA findings, Asteraceae was overrepresented when annual temperatures were very high, and Fabaceae was overrepresented when green space percent was low. At the genus level, *Erigeron* (Asteraceae) was overabundant when annual temperatures were very high, followed by *Helianthus* (Asteraceae), which was also slightly overrepresented in

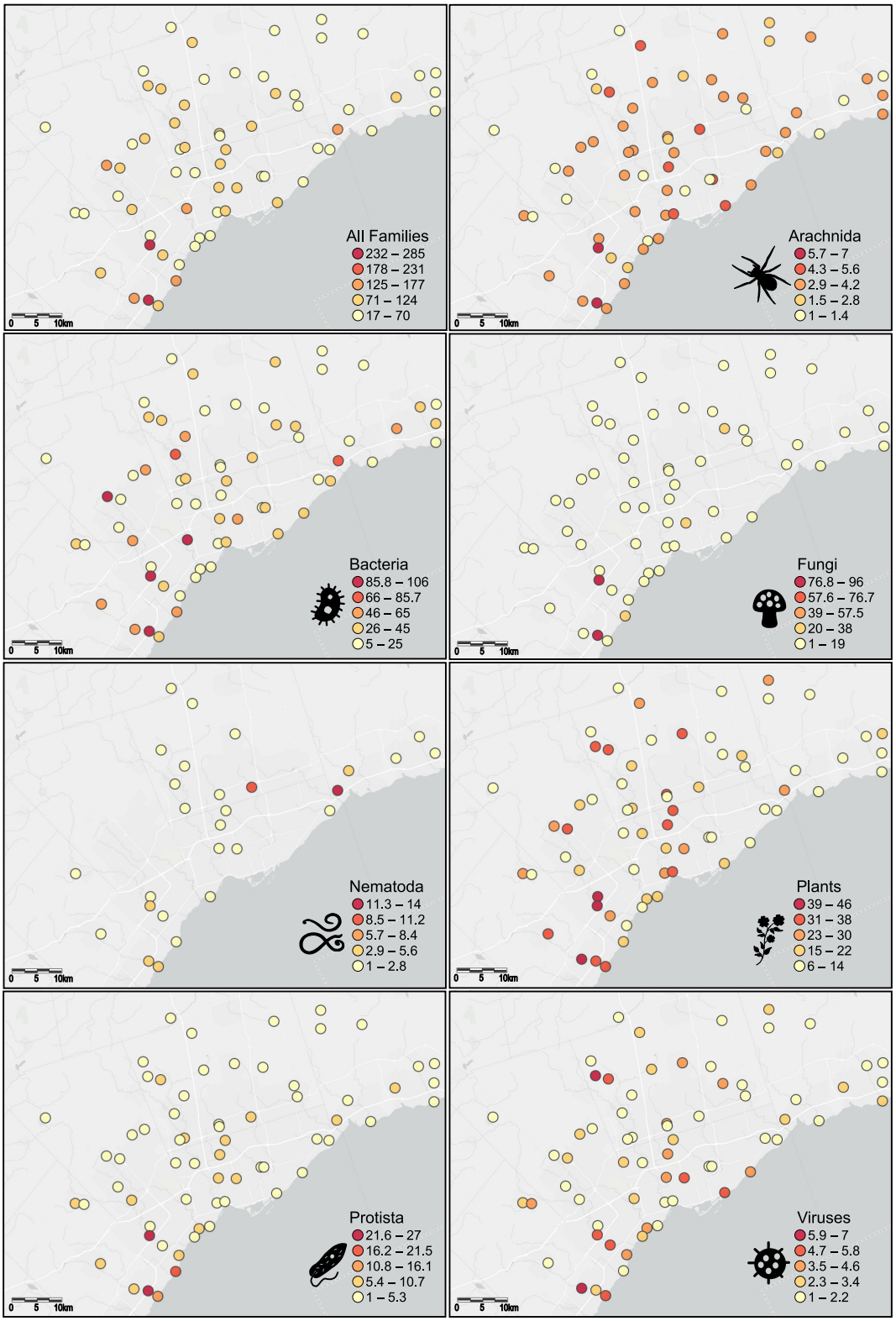

FIGURE 3 Distribution of families across collection sites (circles) for the whole dataset (upper left: all families) and for each distinct group.

very high temperatures (Table S24). SIMPER analyses also revealed several bacterial families and genera driving dissimilarities across categories for each landscape variable. Specifically, Lactobacillaceae drove the greatest dissimilarity between categories for all four landscape variables, with highest representation when temperatures were highest, and when green space, development, and precipitation were low (Table S23). This result was likely driven by the genus *Apilactobacillus* which contributed about 7% dissimilarity

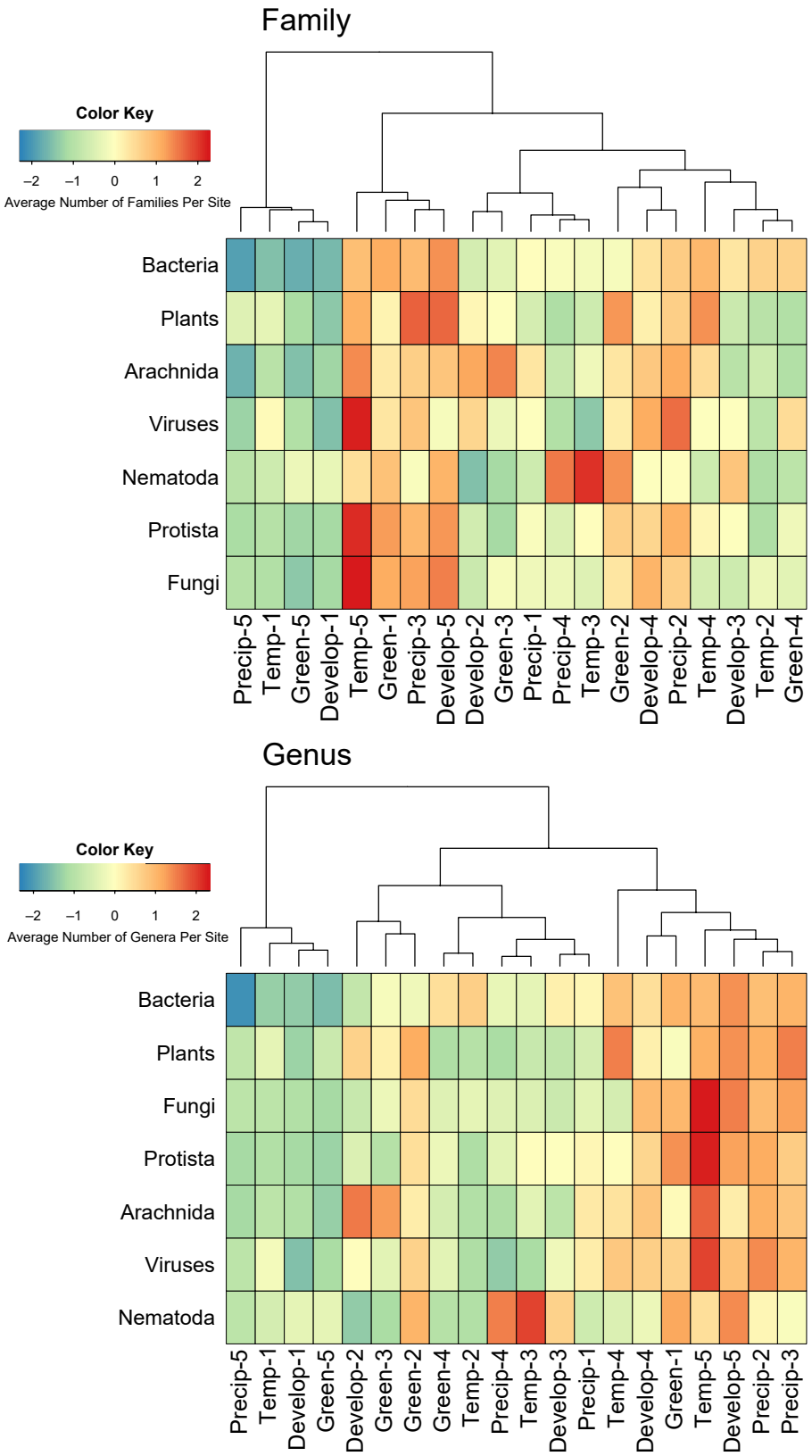

FIGURE 4 Heatmaps of average number of families and genera per site scaled for each environmental variable category (see Table S4). The scale shown is centered and scaled so that mean values are 0 and standard deviations are 1 to visualize domains with high richness with domains with low richness on the same plot. Variables are labelled according to environmental feature: Precip = annual precipitation, Green = green space percentage, Develop = developmental percent, and Temp = annual temperature. 1 = very low, 2 = low, 3 = moderate, 4 = high, and 5 = very high. The heatmap shows 3 or 4 main clusters at the family and genus level, respectively, with the highest average richness per site values forming a distinct cluster in environmental features representative of urban areas.

between landscape categories (Table S24). Enterobacteriaceae was also classified as a top contributor, driving dissimilarities across all four environmental variables with overrepresentation in low annual precipitation (concordant with previous NBDA findings), low annual temperature, moderate development, and moderate green space.

Both GhostKOALA and eggNOG-mapper demonstrated that major functional processes occurring in our non-bee metagenomic taxa are metabolic (e.g., carbon metabolism, biosynthesis of amino acids, and fatty acid metabolism) (Tables S25 and S26, respectively). In addition to metabolism, five other KEGG pathways were identified and include cellular processes, genetic information processing, organismal systems, and environmental information processing. Overall, both methods show that low green space, and high temperatures and development resulted in more reconstructed functional pathways (e.g., cellular processes) alluding to greater taxa productivity in these areas (Figure S13). Contrastingly, high precipitation levels and green space together with low development and a mix of high and low temperature resulted in lower taxa productivity via less reconstructed functional pathways.

3.4 | *Ceratina calcarata* microbiome consists of diverse core taxa

At the family level, the core microbiome (i.e., taxa that occur in ≥50% sites and with ≥1% relative contig abundance) consisted of six bacterial families (Lactobacillaceae, Yersiniaceae, Erwiniaceae, Pseudomonadaceae, Enterobacteriaceae, and Hafniaceae) (Table S11). At the genus level, the microbiome consisted of only three bacterial genera (*Serratia*, *Pseudomonas*, and *Edwardsiella*) (Table S12). Our results identified these families and genera as top contributors for observed differences among categories for precipitation, temperature, development, and green space (Table S23).

Comparison of our *Ceratina* microbiome with other studies on wild bee and non-wild bee microbiomes revealed that the *Ceratina* microbiome is highly diverse with several known core taxa detected, but at low relative abundances (Tables S27 and S28). *Bifidobacterium*, a bacteria involved with digestion of polysaccharides and are core in honey bee microbiomes (Ellegaard et al., 2015; Zheng et al., 2019), was rare with only 0.02% relative contig abundance and found in one site. Similarly, we detected several bacteria that are core in wild and corbiculate bees but were rarely found at relative contig abundances greater than 1% or in more than 50% of the sites in this study. These included *Acinetobacter*, *Sphingomonas*, *Erwinia*, and *Gilliamella*. We detected *Lactobacillus*, *Apilactobacillus*, and *Bacillus*, which are core in corbiculate bees but interestingly were not identified to be a core genus in our study, except at the family level for Lactobacillaceae. *Pseudomonas* is detected in solitary and corbiculate bee microbiomes and was detected at high relative abundance (5.7%) across 33 sites in this study. We also detected *Burkholderia*, a bacterial genus found to be core in other *Ceratina*, across 16 sites with 0.38% relative abundance.

The *C. calcarata* metagenome detected several plant families known to be prevalent in this species' diet. The most detected plant families by overall abundance across the sample region belonged to Asteraceae (19.66% relative contig abundance), followed by Fabaceae (4.50%), Solanaceae (1.71%), Malvaceae (1.68%), and Brassicaceae (1.16%) (Table S11). At the genus level, there were several common plant genera and included daisies (*Erigeron*; 11.87%), sunflowers (*Helianthus*; 2.30%), and clover (*Trifolium*; 1.37%) (Table S12).

Only one viral family was detected with a relative contig abundance above 1% (Myoviridae; 2.28%); but no viral genera produced relative contig abundances above our 1% threshold.

Our metagenomic study also detected several known bee and plant pathogens, although mostly at low abundances, clustering in densely urban sites (Figure 5; Table S29). We detected the arachnid genus *Varroa* (3 sites), and several bacterial genera known to be detrimental to honey and bumble bees such as *Lysinibacillus* (lethal brood disease in stingless bees), *Paenibacillus* (American foulbrood in honey bees), and *Spiroplasma* (neurological disease in honey bees) (Table S29). We also detected the trypansomatid *Crithidia* (parasite of honey and bumble bees) in one densely urban site, and several prominent fungal genera that infect honey bees, including *Ascosphaera* (chalkbrood disease), *Aspergillus* (stonebrood disease), and *Nosema* (spore-forming fungi causing nosemosis disease) all present in sites with high urban development, high temperature and low green space coverage (Tables S8 and S29). Furthermore, we detected several pathogens and parasites of plants, also at low abundances, such as fungal genera *Fusarium* (release of mycotoxins on crop plants), *Alternaria* (brown spot disease), and *Verticillium* (wilt fungus), and the protist *Albugo*, which causes downy mildew in plants (Table S29). The detection of these plant pathogens may reduce the quality and quantity of floral resources required by wild urban bees. We detected the cyst nematode genus *Heterodera* which is known to parasitize a variety of plants. Well known plant pathogenic bacteria such as *Pantoea* was detected across 30 sites with a 6.11% relative contig abundance, and *Clostridium*, an environmental bacterium that can cause plant diseases, was present in 20 sites but at a low relative contig abundance (0.12%) (Table S29). The detection of pathogenic fungi, protists, bacteria, and nematodes across various densely urban sites may enable early detection of bee environmental stressors.

3.5 | Pathogen load and diversity increases as resistance distance increases

We analysed the correlation between population and landscape level parameters against potential bee and plant pathogenicity for 162 sample pairwise comparisons using CoNet. Pathogenicity was not significantly correlated with environmental distance, but was significantly, negatively correlated with relatedness ($R = -0.02$, $p < .05$; Table S30). Although slight, the negative correlation between relatedness and pathogenicity supports that less related individuals harbor more dissimilar pathogens. CoNet analysis also found that relatedness significantly, negatively correlated with resistance distance ($R = -0.14$, $p < .05$; Table S30) and environmental distance ($R = -0.068$, $p < .05$; Table S30), supporting the results of the multiple regression on distance matrices (MRM) analysis. There was one significant,

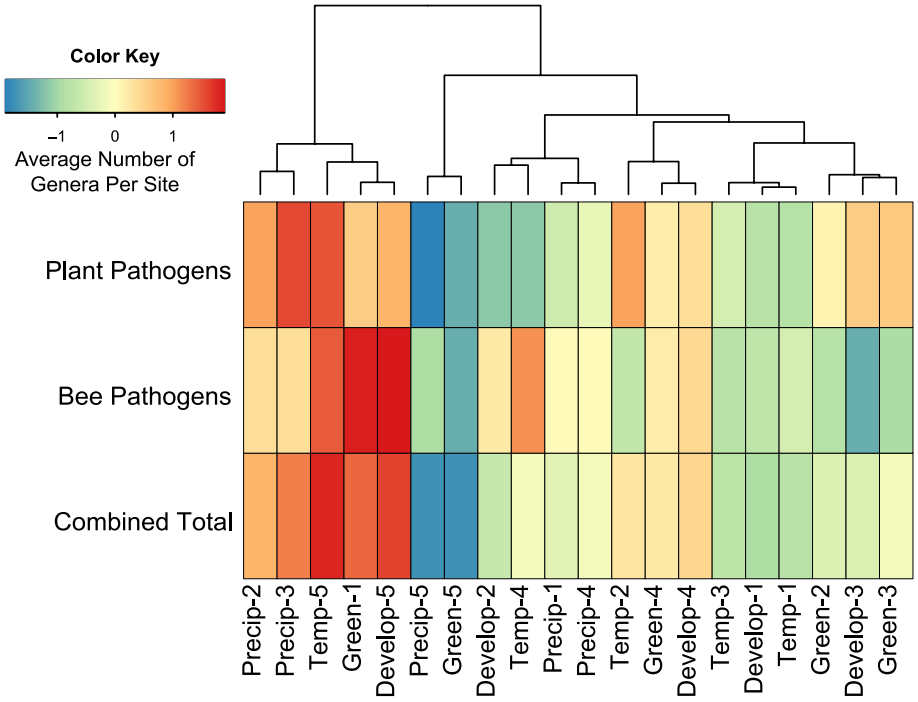

FIGURE 5 Heatmap focused on potential bee ($N = 11$) and plant ($N = 9$) average number of pathogen genera per site for each environmental variable category. Combined total includes both bee and plant pathogen genera counts. Variables are labelled according to environmental feature: Precip = annual precipitation, Green = green space percentage, Develop = developmental percent, and Temp = annual temperature. 1 = very low, 2 = low, 3 = moderate, 4 = high, and 5 = very high. The scale shown is centered and scaled so that mean values are 0 and standard deviations are 1 to visualize domains with high richness with domains with low richness on the same plot. The heatmap emphasizes that high pathogen richness forms a cluster in environmental features representative of urban areas.

positive correlation between pathogenicity and resistance distance ($R = 0.057$, $p < .05$; Table S30). This positive correlation suggests that there is greater pathogen diversity in densely urban areas.

4 | DISCUSSION

We used whole genome sequencing data of *C. calcarata* to conduct population genetics analyses and analyse wild bee metagenome across an urbanization gradient. We assessed *C. calcarata* population structure and connectivity and examined how changes in landscape features affected the abundance, distribution, and presence of bacteria, arachnids, protists, plants, viruses, fungi, and nematodes in female *C. calcarata*. Our findings revealed that *C. calcarata* population in Toronto shows no population structure and high levels of inbreeding, as well as demonstrated the potential importance of vegetation-rich corridors for wild bee dispersal. We also identify how landscape may modify the core microbiome of this bee and detect potential pathogens detrimental to the health and diversity of pollinators and their plant communities.

4.1 | Population structure and connectivity of a wild bee in an urban context

Our population genetics analyses of urban *C. calcarata* revealed one panmictic population, with elevated levels of inbreeding (mean $F_{IS} = 0.083$). The absence of population structure did not support our hypothesis that urban environment contributes to population fragmentation in *C. calcarata*, which was unexpected given the small body size of this bee. Low levels of differentiation across large geographic distances have been previously demonstrated in bumble bees (Kent et al., 2018) and the burrowing bee *Amegilla dawsoni* (Beveridge & Simmons, 2006). The inbreeding value detected in this study aligns with those found in bumble bees in Missouri urban landscapes, with several *Bombus* species revealing F_{IS} values ranging from 0.070 to 0.151 (Austin et al., 2022). However, low genetic diversity does not appear to be unique to urban wild bees. For example, mean F_{IS} was 0.19 ± 0.03 for a temperate bumble bee (*Bombus pascuorum*) and was not significantly impacted by seminatural or agricultural habitats (Herrmann et al., 2007). Another study on Neotropical orchid bees found that mainland urban and rural populations had similar genetic diversity whereas island populations uniquely had reduced genetic diversity (Soro et al., 2017). Our results show that relatedness within the examined population decreased with increased geographic distance (Figure S4). We also demonstrate that resistance scenarios that considered vegetation-rich land use categories (i.e., trees, shrub/scrub, grasses, and crops) as facilitative to bee movement slightly improved the correlation values (Table 1). In contrast when urban surfaces were given low resistance values, the strength of the correlation with relatedness decreased (Table 1). Despite the fact that correlation values were low across all examined hypotheses, changes in correlation values across different scenarios support

our predictions regarding the effects of land use features on gene flow in *C. calcarata*. These findings suggest that dispersal may be driven by vegetation and limited by impervious surfaces, thus re-emphasizing the importance of green spaces for wild bee sustainability in urban landscapes.

4.2 | Taxa diversity is greater in densely urban areas

Overall, taxa family and genera alpha diversity represented by family and genera richness were found to increase in densely urban sites (Figure 3; Figure S7). This agreed with our expectation that taxa diversity would change along the urban gradient. Sites representative of urbanized areas are generally high in temperature and developmental percent, but low in precipitation and green space coverages. Furthermore, when analysing the average number of families and/or genera per site for each environmental category, the highest average richness values formed a cluster in environmental features representative of urban areas (Figure 4). Increased metagenomic diversity in urban sites suggests that wild urban bees are exposed to a greater variety of microbiota. For instance, we found that overall bacterial diversity was highest in urban areas which corroborates the findings of a recent study that also found bacterial richness in *C. calcarata* increased as urbanization increased (Nguyen & Rehan, 2022a). The authors also found that certain core microbiota differed in their abundance depending on urbanization level, such as *Apilactobacillus* and *Acinetobacter* core bacteria overabundant in rural sites, but the fungi *Penicillium*, which plays a role in bee immunity (Yoder et al., 2013), was abundant in urban areas (Nguyen & Rehan, 2022b). Likewise, we also found the change in metagenomic diversity revealed interesting patterns of abundance for core microbiota, such as the core family Lactobacillaceae, which was abundant in sites with high development. These findings and previous studies indicate that urban landscapes can act as refuges (Ayers & Rehan, 2021) for wild bees by facilitating the interaction of diverse and beneficial bacteria but may also increase exposure to potential pathogens and parasites.

4.3 | The core bacterial microbiome

The *C. calcarata* core microbiome harbors a different composition of core microbiota that disagrees with our prediction that the core microbiome of our urban bee would differ from that of previous bee studies. Six bacterial families were identified as core in *C. calcarata*. These included prominent bacterial families such as Lactobacillaceae, which has been recognized to play an important role in bee immunity and carbohydrate metabolism (Daisley et al., 2020; Kwong et al., 2017; Lang et al., 2022; Voulgari-Kokota et al., 2019). Enterobacteriaceae, which are facultative anaerobes involved in bee metabolism by breaking down sugars and nitrogen (Anderson et al., 2011), was the only bacterial family to be overrepresented when annual precipitation was low. Furthermore, network analyses revealed several hub genera belonging to Enterobacteriaceae clustering with plants, suggesting that bees foraging in drier areas may be exposed to higher occurrences of Enterobacteriaceae that occur naturally on the plant phyllosphere (Aydogan et al., 2018; Cernava et al., 2019).

We also detected Yersiniaceae, Erwiniaceae, Pseudomonadaceae, and Hafniaceae as core families, and at the genus level, *Serratia* (Yersiniaceae), *Pseudomonas* (Pseudomonadaceae), and *Edwardsiella* (Hafniacea). While *Pseudomonas* has been detected as core in solitary and corbiculate bees (De Paula et al., 2021; Donkersley et al., 2018; Graystock et al., 2017), *Serratia* and *Edwardsiella* have been detected in honey bee microbiomes but are not considered core (El Khoury et al., 2022; Nogrado et al., 2019; Raymann et al., 2018). However, *Serratia* and *Edwardsiella* are classified as environmental bacteria and were found to increase in abundance in honey bees exposed to agricultural pesticides, which are more common in cities near agricultural spaces (Mahé et al., 2021; Nogrado et al., 2019). While the *Ceratina* microbiome shares some similarities to other bees, solitary bee microbiomes are heavily influenced by the environment, and urban populations may be prime locations to test for microbiome shifts as a response to environmental change and contamination.

4.4 | Pathogens linked to urban land use

Metagenomic analysis is useful for identifying key players in bee microbiomes and detection of potential pathogens to both the bee and plants in their environment. Sites displaying symptoms of the urban heat island effect (e.g., high temperatures in densely urban areas, Oke, 1973), demonstrated highest alpha diversity across domains (Figures 3 and 4) and across pathogens only (Figure 5). Higher taxa diversity in urban landscapes poses the risk of greater exposure to potential pathogens. Indeed, our results agreed with our prediction that pathogen load and diversity would be greater in urban landscapes. Among these potential pathogens, *Aspergillus* was relatively common and found across over 30% (20 sites) of the landscape assayed. *Aspergillus* is a genus with several species pathogenic to almost all organisms from mammals to insects and causes stonebrood disease in honey bee larvae (Foley et al., 2014; Seyedmousavi et al., 2015). Direct effects of *Aspergillus* on solitary bee larvae are currently unknown but is known to attack brood cells of solitary, ground-nesting alkali bees (*Nomia*; Batra & Bohart, 1969) and increased mortality in the solitary chimney bee (*Diadasia bituberculata*) by causing pollen provision spoilage (Linsley & MacSwain, 1952). We also detected the parasitic arachnid *Varroa* (3 sites), which transmits deformed wing virus and acute bee paralysis virus in honey bees (Noël et al., 2020; Traynor et al., 2020). Non-*Apis* insects visiting flowers previously visited by an infected honey bee are susceptible to infections of deformed wing virus due to pathogen spill over between bee species (Burnham et al., 2021; Santamaria et al., 2018; Tehel et al., 2016). The well-studied honey bee bacterial pathogen *Spiroplasma* was detected across 15 sites, most of which are densely urban. *Spiroplasma* spp. have been detected in bumble bees and mason

bees, but it is unclear what the effects are in non-*Apis* bee species (Fünfhaus et al., 2018; Macfarlane et al., 1995).

Our metagenomic data revealed *C. calcarata* to be associated with plants from Asteraceae (asters), Brassicaceae (mustards), Fabaceae (legumes), Solanaceae (nightshades), and Malvaceae (mallows). These plant families have important links to wild bee diet, for example, Fabaceae (legumes) are rich in essential proteins needed for healthy larvae development (Cole et al., 2022), and Asteraceae (asters) boasts the largest richness of ornamental flowers despite being relatively protein poor (Spear et al., 2016). Furthermore, these plant families were mainly abundant in urban sites and thus act as important food sources for urban bees in addition to promoting diverse ecosystem services (Gamfeldt et al., 2013; Isbell et al., 2011; Keesing et al., 2010). However, high plant diversity, especially with the rise of exotic ornamental plants, might in fact pose more of a risk to wild native bees by harboring increased diversity of pathogens and parasites (Figure 5; Figueroa et al., 2019; Graystock et al., 2015). For instance, while most potential pathogens and parasites were found in a few sites in our study, *Pantoea* (Erwiniaceae) was abundant across 30 sites, mainly in moderate to highly developed areas. *Pantoea* is a highly diverse bacterial genus that can form mutualistic relationships with insects but harbors many pathogenic species for plants (Walterson & Stavrinides, 2015). Other taxa may act as either beneficial or pathogenic depending on environmental conditions or species. For example, *Clostridium*, an anaerobic, environmental bacterium that can promote plant growth (Figueiredo et al., 2020; Polyanskaya et al., 2002) but can also cause several plant diseases when soils are flooded (Figueiredo et al., 2020; Spigaglia et al., 2020), was prevalent across 20 sites also mainly from moderately to highly developed sites. Additionally, the plant metagenomic data highlights that while *C. calcarata* are classified as polylectic, the majority of the foraging known for this bee genus is more broadly focused on a few plants (Dew et al., 2020; Kobayashi-Kidokoro & Higashi, 2010; Lawson et al., 2016, 2020; McFrederick & Rehan, 2016, 2019). This emphasizes the need to monitor the abundance and distribution of these plants in cities as important foraging sources, as well as potential hubs for disease spread, will be crucial for wild bees that display flower constancy in urban environments.

Overall, we found that urban bees harbor a microbiome that appears more productive, as they had a greater number of reconstructed functional pathways than their rural conspecifics (Figure S13). We speculate that this is potentially due to urban bees interacting with greater taxa loads and diversity found on plants, which were more abundant in urban sites, and a general trend of greater bacterial diversity in urban sites (Figure 3; Figure S7). Greater bacterial diversity was also found in urban *C. calcarata* from a recent study (Nguyen & Rehan, 2022a). Urban bees also run an increased risk of encountering more harmful taxa. Indeed, alpha diversity of bee and plant pathogens detected in this study were highest in sites with very high temperatures and development (Figure 5), which is consistent with studies suggesting that urbanized landscapes increase pathogen and parasite diversity (Youngsteadt et al., 2015). Indeed, when comparing pathogenicity with population and landscape parameters across pairwise samples, pathogenicity was significantly positively correlated with resistance distance (Table S28), further supporting the expectation that as resistance between sites increases—reflective of urban systems—pathogen diversity increases. Finally, we observed that as relatedness between individuals decreased, pathogenicity increased. Urban landscapes are often fragmented, and thus are more likely to promote differentiation, exposing vulnerable wild bee populations to diverse pathogens. The intricate interplay of plant-pollinator systems shows that both can act as hubs and vectors of diseases for each other. Our findings provide important insights into the diversity of wild bee responses and adaptations to city landscapes. Each species is sure to have its own response thresholds to environmental stressors, such as susceptibility to disease vectors.

4.5 | Conclusion

In summary, we characterized the metagenome of a wild bee species across an urbanization gradient, revealing this species' consistent core microbiome and floral associations. This study reveals a highly diverse metagenome with taxa detected across seven domains. Increased green spaces, notably shrubs and grasses, are required to maintain high connectedness between nesting and foraging sites for urban stem nesting bees. While low genetic diversity appears to be recurrent in rural and urban bees, low genetic diversity in urban bees may be more detrimental where gene flow is limited due to a fragmented landscape and susceptibility to genetic bottlenecks is more likely as urbanization increases. Future studies are needed to explore the link between reduced heterozygosity and the fitness of wild bees in urban landscapes. We identified several known bee and plant pathogens that occur in densely urban sites, revealing the potential for early detection and monitoring of threats to wildlife in cities. Conservation efforts would benefit from the integration of population genomics with metagenomics. This combined whole genome sequencing approach provides a comprehensive tool to assess overall health of wild bees in urban settings but can also be applied across a broad range of wildlife and landscapes more generally.

AUTHOR CONTRIBUTIONS

Sandra Rehan conceived the experiment. Evan Kelemen conducted field sampling and DNA extractions. Katherine Chau and Farida Samad-zada performed the bioinformatics analyses for metagenomic and population genetics, respectively. All authors were involved with analysis and interpretation of data, as well as writing of the manuscript. All authors were involved with manuscript revisions and approved the final manuscript.

ACKNOWLEDGMENTS

We thank members of the Rehan lab for constructive feedback on earlier versions of this manuscript. Thanks to Alessia Schembri, Phuong Nguyen, and Yousaf Malik for assistance with field collection of bee specimens. Analysis of all data was enabled in part by support provided by the high-performance computing (HPC) clusters from

Compute Canada, Compute Ontario and SHARCNET. We thank Jose Sergio Hleap for assistance with the HPC clusters. This work was funded by Mitacs Elevate Postdoctoral Fellowship to KDC, NSF Biological Collections Postdoctoral Fellowship DBI-1906494 to EPK, Foundation for Food and Agricultural Research Pollinator Health fund 549038, Weston Family Foundation Microbiome Initiative, NSERC Discovery Grant, Supplement and E.W.R Steacie Memorial Fellowship to SMR.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.7884663. Sequence data are publicly available in NCBI, BioProject: PRJNA725238. Scripts and bioinformatics pipeline are available at https://github.com/kdbchau/Bee-Metagenomics.

ORCID

Evan P. Kelemen https://orcid.org/0000-0003-3600-9878
Sandra M. Rehan https://orcid.org/0000-0002-6441-5155

REFERENCES

Adriaensen, F., Chardon, J. P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., & Matthysen, E. (2003). The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning*, *64*(4), 233–247. https://doi.org/10.1016/S0169-2046(02)00242-6

Aizen, M. A., Garibaldi, L. A., Cunningham, S. A., & Klein, A. M. (2009). How much does agriculture depend on pollinators? Lessons from long-term trends in crop production. *Annals of Botany*, *103*(9), 1579–1588. https://doi.org/10.1093/aob/mcp076

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson, K., Sheehan, T., Eckholm, B., Mott, B., & DeGrandi-Hoffman, G. (2011). An emerging paradigm of colony health: Microbial balance of the honey bee and hive (*Apis mellifera*). *Insectes Sociaux*, *58*, 431–444. https://doi.org/10.1007/s00040-011-0194-6

Austin, M. W., Tripodi, A. D., Strange, J. P., & Dunlap, A. S. (2022). Bumble bees exhibit body size clines across an urban gradient despite low genetic differentiation. *Scientific Reports*, *12*, 4166. https://doi.org/10.1038/s41598-022-08093-4

Aydogan, E. L., Moser, G., Müller, C., Kämpfer, P., & Glaeser, S. P. (2018). Long-term warming shifts the composition of bacterial communities in the Phyllosphere of *Galium album* in a permanent grassland field-experiment. *Frontiers in Microbiology*, *9*, 144. https://doi.org/10.3389/fmicb.2018.00144

Ayers, A. C., & Rehan, S. M. (2021). Supporting bees in cities: How bees are influenced by local and landscape features. *Insects*, *12*(2), 128. https://doi.org/10.3390/insects12020128

Ballare, K. M., & Jha, S. (2021). Genetic structure across urban and agricultural landscapes reveals evidence of resource specialization and philopatry in the eastern carpenter bee, *Xylocopa Virginica* L. *Evolutionary Applications*, *14*(1), 136–149. https://doi.org/10.1111/eva.13078

Banaszak-Cibicka, W., Twerd, L., Fliszkiewicz, M., Giejdasz, K., & Langowska, A. (2018). City parks vs. natural areas—Is it possible to preserve a natural level of bee richness and abundance in a city park? *Urban Ecosystem*, *21*, 599–613. https://doi.org/10.1007/s11252-018-0756-8

Bartomeus, I., Ascher, J. S., Gibbs, J., Danforth, B. N., Wagner, D. L., Hedtke, S. M., & Winfree, R. (2013). Historical changes in northeastern US bee pollinators related to shared ecological traits. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(12), 4656–4660. https://doi.org/10.1073/pnas.1218503110

Batra, S. W. T. (1984). Solitary bees. *Scientific American*, *250*(2), 120–127. https://doi.org/10.1038/scientificamerican0284-120

Batra, S. W. T., & Bohart, G. E. (1969). Alkali bees: Response of adults to pathogenic fungi in brood cells. *Science*, *165*(3893), 607. https://doi.org/10.1126/science.165.3893.607

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (2004). *GENETIX 4.05, windows TM software for population genetics*. Laboratoire génome, Populations, Interactions, CNRS UMR.

Beveridge, M., & Simmons, L. W. (2006). Panmixia: An example from Dawson's burrowing bee (*Amegilla dawsoni*) (hymenoptera: Anthophorini): Panmixia: An example from *Amegilla Dawsoni*. *Molecular Ecology*, *15*(4), 951–957. https://doi.org/10.1111/j.1365-294X.2006.02846.x

Breeze, T. D., Bailey, A. P., Balcombe, K. G., & Potts, S. G. (2011). Pollination services in the UK: How important are honeybees? *Agriculture, Ecosystems & Environment*, *142*(3–4), 137–143. https://doi.org/10.1016/j.agee.2011.03.020

Burdine, J. D., & McCluney, K. E. (2019). Differential sensitivity of bees to urbanization-driven changes in body temperature and water content. *Scientific Reports*, *9*, 1643. https://doi.org/10.1038/s41598-018-38338-0

Burnham, P. A., Alger, S. A., Case, B., Boncristiani, H., Hébert-Dufresne, L., & Brody, A. K. (2021). Flowers as dirty doorknobs: Deformed wing virus transmitted between *Apis mellifera* and *Bombus impatiens* through shared flowers. *Journal of Applied Ecology*, *58*(10), 2065–2074. https://doi.org/10.1111/1365-2664.13962

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper V2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, *38*(12), 582–5829. https://doi.org/10.1093/molbev/msab293

Cardoso, M. C., & Gonçalves, R. B. (2018). Reduction by half: The impact on bees of 34 years of urbanization. *Urban Ecosystem*, *21*, 943–949. https://doi.org/10.1007/s11252-018-0773-7

Cernava, T., Erlacher, A., Soh, J., Sensen, C. W., Grube, M., & Berg, G. (2019). Enterobacteriaceae dominate the core microbiome and contribute to the resistome of arugula (*Eruca sativa* mill.). *Microbiome*, *7*, 13. https://doi.org/10.1186/s40168-019-0624-7

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ. Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Cohen, H., McFrederick, Q. S., & Philpott, S. M. (2020). Environment shapes the microbiome of the blue orchard bee, Osmia lignaria. *Microbial Ecology*, *80*(4), 897–907. https://doi.org/10.1007/s00248-020-01549-y

Cole, L. J., Baddeley, J. A., Robertson, D., Topp, C. F. E., Walker, R. L., & Watson, C. A. (2022). Supporting wild pollinators in agricultural landscapes through targeted legume mixtures. *Agriculture, Ecosystems & Environment*, *323*, 107648. https://doi.org/10.1016/j.agee.2021.107648

Daisley, B. A., Pitek, A. P., Chmiel, J. A., Gibbons, S., Chernyshova, A. M., Al, K. F., Faragalla, K. M., Burton, J. P., Thompson, G. J., & Reid, G.

(2020). *Lactobacillus* spp. attenuate antibiotic-induced immune and microbiota dysregulation in honey bees. *Communications Biology*, *3*, 534. https://doi.org/10.1038/s42003-020-01259-8

Dalapicolla, J., Prado, J. R., Percequillo, A. R., & Knowles, L. L. (2021). Functional connectivity in sympatric spiny rats reflects different dimensions of Amazonian forest-association. *Journal of Biogeography*, *48*(12), 3196–3209. https://doi.org/10.1111/jbi.14281

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danforth, B. N., Minckley, R. L., Neff, J. L., & Fawcett, F. (2019). *The solitary bees: Biology, evolution, conservation*. Princeton University Press.

De Paula, G. T., Menezes, C., Pupo, M. T., & Rosa, C. A. (2021). Stingless bees and microbial interactions. *Current Opinion in Insect Science*, *44*, 41–47. https://doi.org/10.1016/j.cois.2020.11.006

Des Roches, S., Brans, K. I., Lambert, M. R., Rivkin, R., Savage, A. M., Schell, C. J., Correa, C., De Meester, L., Diamond, S. E., Grimm, N. B., Harris, N. C., Govaert, L., Hendry, A. P., Johnson, M. T. J., Munshi-South, J., Palkovacs, E. P., Szulkin, M., Urban, M. C., Verrelli, B. C., & Alberti, M. (2020). Socio-eco-evolutionary dynamics in cities. *Evolutionary Applications*, *14*(1), 248–267. https://doi.org/10.1111/eva.13065

Dew, R. M., McFrederick, Q. S., & Rehan, S. M. (2020). Diverse diets with consistent Core microbiome in wild bee pollen provisions. *Insects*, *11*(8), 499. https://doi.org/10.3390/insects11080499

Dhami, I., Arano, K. G., Warner, T. A., Gazal, R. M., & Joshi, S. (2011). Phenology of trees and urbanization: A comparative study between new York City and Ithaca, New York. *Geocarto International*, *26*(7), 507–526. https://doi.org/10.1080/10106049.2011.607517

Dharampal, P. S., Carlson, C., Currie, C. R., & Steffan, S. A. (2019). Pollen-borne microbes shape bee fitness. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1904), 20182894. https://doi.org/10.1098/rspb.2018.2894

Donkersley, P., Rhodes, G., Pickup, R. W., Jones, K. C., & Wilson, K. (2018). Bacterial communities associated with honeybee food stores are correlated with land use. *Ecology and Evolution*, *8*(10), 4743–4756. https://doi.org/10.1002/ece3.3999

Durrer, S., & Schmid-Hempel, P. (1994). Shared use of flowers leads to horizontal pathogen transmission. *Proceedings of the Royal Society B: Biological Sciences*, *258*(1353), 299–302. https://doi.org/10.1098/rspb.1994.0176

El Khoury, S., Giovenazzo, P., & Derome, N. (2022). Endogenous honeybee gut microbiota metabolize the pesticide clothianidin. *Microorganisms*, *10*(3), 493. https://doi.org/10.3390/microorganisms10030493

Ellegaard, K. M., Tamarit, D., Javelind, E., Olofsson, T. C., Anderrson, S. G. E., & Vásquez, A. (2015). Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics*, *16*, 284. https://doi.org/10.1186/s12864-015-1476-6

Elmgvist, T., Andersson, E., McPhearson, T., Bai, X., Bettencourt, L., Brondizio, E., Colding, J., Daily, G., Folke, C., Grimm, N., Haase, D., Ospina, D., Parnell, S., Polasky, S., Seto, K. C., & Van Der Leeuw, S. (2021). Urbanization in and for the Anthropocene. *npj Urban Sustainability*, *1*, 6. https://doi.org/10.1038/s42949-021-00018-w

Engel, P., Kwong, W. K., McFrederick, Q., Anderson, K. E., Barribeau, S. M., Chandler, J. A., Cornman, R. S., Dainat, J., de Miranda, J. R., Doublet, V., Emery, O., Evans, J. D., Farinelli, L., Flenniken, M. L., Granberd, F., Grasis, J. A., Gauthier, L., Hayer, J., Koch, H., … Dainat, B. (2016). The bee microbiome: Impact on bee health and model for evolution and ecology of host-microbiome interactions. *American Society for Microbiology*, *7*(2), 1–9. https://doi.org/10.1128/mBio.02164-15

Engel, P., Martinson, V. G., & Moran, N. A. (2012). Functional diversity within the simple gut microbiota of the honey bee. *PNAS*, *109*(27), 11002–11007. https://doi.org/10.1073/pnas.1202970109

Faust, K., & Raes, J. (2016). CoNet app: Inference of biological association networks using Cytoscape. *F1000Research*, *5*, 1519. https://doi.org/10.12688/f1000research.9050.2

Figueiredo, G. G. O., Lopes, V. R., Romano, T., & Camara, M. C. (2020). Clostridium. In N. Amaresan, M. S. Kumar, K. Annapurna, K. Kumar, & A. Sankaranarayanan (Eds.), *Beneficial microbes in agro-ecology* (pp. 477–491). Academic Press.

Figueroa, L. L., Blinder, M., Grincavitch, C., Jelink, A., Mann, E. K., Merva, L. A., Metz, L. E., Zhao, A. Y., Irwin, R. E., McArt, S. H., & Adler, L. S. (2019). Bee pathogen transmission dynamics: Deposition, persistence and acquisition on flowers. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1903), 20190603. https://doi.org/10.1098/rspb.2019.0603

Fitch, G., Wilson, C. J., Glaum, P., Vaidya, C., Simao, M.-C., & Jamieson, M. A. (2019). Does urbanization favour exotic bee species? Implications for the conservation of native bees in cities. *Biology Letters*, *15*(12), 20190574. https://doi.org/10.1098/rsbl.2019.0574

Foley, K., Fazio, G., Jensen, A. B., & Hughes, W. O. H. (2014). The distribution of *aspergillus spp*. opportunistic parasites in hives and their pathogenicity to honey bees. *Veterinary Microbiology*, *169*, 203–210. https://doi.org/10.1016/j.vetmic.2013.11.029

Freitas, F. V., Branstetter, M. G., Griswold, T., & Almeida, E. A. J. M. B. (2020). Partitioned gene-tree analyses and gene-based topology testing help resolve incongruence in a phylogenomic study of host-specialist bees (Apidae: Eucerinae). *Molecular Biology and Evolution*, *38*(3), 1090–1100. https://doi.org/10.1093/molbev/msaa277

Fünfhaus, A., Ebeling, J., & Genersch, E. (2018). Bacterial pathogens of bees. *Current Opinion in Insect Science*, *26*, 89–96. https://doi.org/10.1016/j.cois.2018.02.008

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M. C., Fröberg, M., Stendahl, J., Philipson, C. D., Mikusiński, G., Andersson, E., Westerlund, B., Andrén, H., Moberg, F., Moen, J., & Bengtsson, J. (2013). Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications*, *4*, 1340. https://doi.org/10.1038/ncomms2328

Geslin, B., Gachet, S., Deschamps-Cottin, M., Flacher, F., Ignace, B., Knoploch, C., Meineri, É., Robles, C., Ropars, L., Schurr, L., & Le Féon, V. (2020). Bee hotels host a high abundance of exotic bees in an urban context. *Acta Oecologica*, *105*, 103556. https://doi.org/10.1016/j.actao.2020.103556

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*(7), 1–19. https://doi.org/10.18637/jss.v022.i07

Graystock, P., Goulson, D., & Hughes, W. O. H. (2015). Parasites in bloom: Flowers aid dispersal and transmission of pollinator parasites within and between bee species. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1813), 20151371. https://doi.org/10.1098/rspb.2015.1371

Graystock, P., Ng, W. H., Parks, K., Tripodi, A. D., Muñiz, P. A., Fersch, A. A., Myers, C. R., McFrederick, W. S., & McArt, S. H. (2020). Dominant bee species and floral abundance drive parasite temporal dynamics in plant-pollinator communities. *Nature Ecology & Evolution*, *4*, 1358–1367. https://doi.org/10.1038/s41559-020-1247-x

Graystock, P., Rehan, S. M., & McFrederick, Q. S. (2017). Hunting for healthy microbiomes: Determining the core microbiomes of *Ceratina*, *Megalopta*, and *Apis* bees and how they associate with microbes in bee collected pollen. *Conservation Genetics*, *18*, 701–711. https://doi.org/10.1007/s10592-017-0937-7

Guo, S., Xiong, W., Hang, X., Gao, Z., Jiao, Z., Liu, H., Mo, Y., Zhang, N., Kowalchuk, G. A., Shen, Q., & Geisen, S. (2021). Protists as main indicators and determinants of plant performance. *Microbiome*, *9*, 64. https://doi.org/10.1186/s40168-021-01025-w

Hall, K. R., Anantharaman, R., Landau, V. A., Clark, M., Dickson, B. G., Jones, A., Platt, J., Edelman, A., & Shah, V. B. (2021). Circuitscape in Julia: Empowering dynamic approaches to connectivity assessment. *Land*, *10*(3), 301. https://doi.org/10.3390/land10030301

Hamblin, A. L., Youngsteadt, E., & Frank, S. D. (2018). Wild bee abundance declines with urban warming, regardless of floral density. *Urban Ecosystem*, *21*, 419–428. https://doi.org/10.1007/s11252-018-0731-4

Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontological Electronica*, *4*(1), 1–9.

Hanula, J. L., Horn, S., & O'Brien, J. J. (2015). Have changing forests conditions contributed to pollinator decline in the southeastern United States? *Forest Ecology and Management*, *348*, 142–152. https://doi.org/10.1016/j.foreco.2015.03.044

Herrmann, F., Westphal, C., Mortiz, R. F., & Stefan-Dewenter, I. (2007). Genetic diversity and mass resources promote colony size and forager densities of a social bee (*Bombus pascuorum*) in agricultural landscapes. *Molecular Ecology*, *8*(1), 1–14. https://doi.org/10.1111/j.1365-294X.2007.03226.x

Hijmans, R. J., Williams, E., Vennes, C., & Hijmans, M. R. J. (2019). Package 'geosphere'. *Spherical Trigonometry*, 1–5.

Hugenholtz, P., & Tyson, G. W. (2008). Metagenomics. *Nature*, *455*, 481–483. https://doi.org/10.1038/455481a

Hülsmann, M., von Wehrden, H., Klein, A.-M., & Leonhardt, S. D. (2015). Plant diversity and composition compensate for negative effects of urbanization on foraging bumble bees. *Apidologie*, *46*, 760–770. https://doi.org/10.1007/s13592-015-0366-x

Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W. S., Reich, P. B., Scherer-Lorenzen, M., Schmid, B., Tilman, D., van Ruijven, J., Weigelt, A., Wilsey, B. J., Zavaleta, E. S., & Loreau, M. (2011). High plant diversity is needed to maintain ecosystem services. *Nature*, *477*, 199–202. https://doi.org/10.1038/nature10282

Jaffé, R., Veiga, J. C., Pope, N. S., Lanes, É. C. M., Carvalho, C. S., Alves, R., Andrade, S. C. S., Arias, M. C., Bonatti, V., Carvalho, A. T., de Castro, M. S., Contrera, F. A. L., Francoy, T. M., Freitas, B. M., Giannini, T. C., Hrncir, M., Martins, C. F., Oliveira, G., Saraiva, A. M., … Imperatriz-Fonseca, V. L. (2019). Landscape genomics to the rescue of a tropical bee threatened by habitat loss and climate change. *Evolutionary Applications*, *12*(6), 1164–1177. https://doi.org/10.1111/eva.12794

Jha, S. (2015). Contemporary human-altered landscapes and oceanic barriers reduce bumble bee gene flow. *Molecular Ecology*, *24*(5), 993–1006. https://doi.org/10.1111/mec.13090

Jha, S., & Kremen, C. (2013). Urban land use limits regional bumble bee gene flow. *Molecular Ecology*, *22*(9), 2483–2495. https://doi.org/10.1111/mec.12275

Kabano, P., Lindley, S., & Harris, A. (2021). Evidence of urban heat Island impacts on the vegetation growing season length in a tropical city. *Landscape and Urban Planning*, *206*, 103989. https://doi.org/10.1016/j.landurbplan.2020.103989

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kapheim, K. M., Johnson, M. M., & Jolley, M. (2021). Composition and acquisition of the microbiome in solitary, ground-nesting alkali bees. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-82573-x

Kapheim, K. M., Rao, V. D., Yeoman, C. J., Wilson, B. A., White, B. A., Goldenfeld, N., & Robinson, G. E. (2015). Caste-specific differences in hindgut microbial communities of honey bees (*Apis mellifera*). *PLoS One*, *10*(4), e0123911. https://doi.org/10.1371/journal.pone.0123911

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P. (2021). Global land use/land cover with Sentinel 2 and deep learning. *IEEE International Geoscience and Remote Sensing Symposiumm IGARSS*. 4704–4707. https://doi.org/10.1109/IGARSS47720.2021.9553499

Keesing, F., Belden, L. K., Daszak, P., Dobson, A., Harvell, C. D., Holt, R. D., Hudson, P., Jolles, A., Jones, K. E., Mitchell, C. E., Myers, S. S., Bogich, T., & Ostfeld, R. S. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, *468*, 647–652. https://doi.org/10.1038/nature09575

Kelemen, E. P., & Rehan, S. M. (2021). Opposing pressures of climate and land-use change on a native bee. *Global Change Biology*, *27*(5), 1017–1026. https://doi.org/10.1111/gcb.15468

Kent, C. F., Dey, A., Patel, H., Tsvetkov, N., Tiwari, T., MacPhail, V. J., Gobeil, Y., Harpur, B. A., Gurtowski, J., Schatz, M. C., Colla, S. R., & Zayed, A. (2018). Conservation genomics of the declining north American bumblebee *Bombus terricola* reveals inbreeding and selection on immune genes. *Frontiers in Genetics*, *9*, 316. https://doi.org/10.3389/fgene.2018.00316

Kleijn, D., Winfree, R., Bartomeus, I., Carvalheiro, L. G., Henry, M., Isaacs, R., Klein, A.-M., Kremen, C., M'Gonigle, L. K., Rader, R., Ricketts, T. H., Williams, N. M., Adamson, N. L., Ascher, J. S., Báldi, A., Batáry, P., Benjamin, F., Biesmeijer, J. C., Blitzer, E. J., … Potts, S. G. (2015). Delivery of crop pollination services is an insufficient argument for wild pollinator conservation. *Nature Communications*, *6*, 7414. https://doi.org/10.1038/ncomms8414

Klein, A.-M., Vaissière, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., & Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1608). https://doi.org/10.1098/rspb.2006.3721

Kobayashi-Kidokoro, M., & Higashi, S. (2010). *Flower constancy in the generalist pollinator Ceratina flavipes (Hymenoptera: Apidae): An evaluation by pollen analysis* (p. 2010). A Journal of Entomology. https://doi.org/10.1155/2010/891906

Kwong, W. K., Engel, P., Koch, H., & Moran, N. A. (2014). Genomics and host specialization of honey bee and bumble bee gut symbionts. *PNAS*, *111*(31), 11509–11514. https://doi.org/10.1073/pnas.1405838111

Kwong, W. K., Medina, L. A., Kock, H., Sing, K.-W., Soh, E. J. Y., Ascher, J. S., Jaffé, R., & Moran, N. A. (2017). Dynamic microbiome evolution in social bees. *Science Advances*, *3*(3), 1–17. https://doi.org/10.1126/sciadv.1600513

Lang, H., Duan, H., Wang, J., Zhang, W., Guo, J., Zhang, X., Hu, X., & Zheng, H. (2022). Specific strains of honeybee gut *lactobacillus* stimulate host immune system to protect against pathogenic *hafnia alvei*. *American Society for Microbiology*, *10*(1), e0189621. https://doi.org/10.1128/spectrum.01896-21

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559. https://doi.org/10.1186/1471-2105-9-559

Lawson, S. P., Ciaccio, K. N., & Rehan, S. M. (2016). Maternal manipulation of pollen provisions affects worker production in a small carpenter bee. *Behavioral Ecology and Sociobiology*, *70*, 1891–1900. https://doi.org/10.1007/s00265-016-2194-z

Lawson, S. P., Kennedy, K. B., & Rehan, S. M. (2020). Pollen composition significantly impacts the development and survival of the native small carpenter bee, *Ceratina calcarata*. *Ecological Entomology*, *46*(2), 232–239. https://doi.org/10.1111/een.12955

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, L., Solvi, C., Zhang, F., Qi, Z., Chittka, L., & Zhao, W. (2021). Gut microbiome drives individual memory variation in bumblebees. *Nature Communications*, *12*, 6588. https://doi.org/10.1038/s41467-021-26833-4

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Linsley, E. G., & MacSwain, J. W. (1952). Notes on some effects of parasitism upon a small population of *Diadasia bituberculata* (Cresson). *Pan-Pacific Entomologist*, *28*, 131–135 https://biostor.org/reference/225390

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 500. https://doi.org/10.1186/s13059-014-0550-8

MacFarlane, R. P., Lipa, J. J., & Liu, H. J. (1995). Bumble bee pathogens and internal enemies. *Bee World*, *76*, 130–148. https://doi.org/10.1080/0005772X.1995.11099259

Maechler, M. (2022). sfsmisc: utilities from 'Seminar fuer Statistik' ETH Zurich. R package version 1.1.-14.

Magnacca, K. N., & Brown, M. J. F. (2012). DNA barcoding a regional fauna: Irish solitary bees. *Molecular Ecology Resources*, *12*(6), 990–998. https://doi.org/10.1111/1755-0998.12001

Mahé, C., Jumarie, C., & Boily, M. (2021). The countryside or the city: Which environment is better for the honeybee? *Environmental Research*, *195*, 110784. https://doi.org/10.1016/j.envres.2021.110784

Mathiasson, M. E., & Rehan, S. M. (2019). Status changes in the wild bees of North-Eastern North America over 125 years revealed through museum specimens. *Insect Conservation and Diversity*, *12*, 278–288. https://doi.org/10.1111/icad.12347

McFrederick, Q. S., & Rehan, S. M. (2016). Characterization of pollen and bacterial community composition in brood provisions of a small carpenter bee. *Molecular Ecology*, *25*(10), 2302–2311. https://doi.org/10.1111/mec.13608

McFrederick, Q. S., & Rehan, S. M. (2019). Wild bee pollen usage and microbial communities Co-vary across landscapes. *Microbial Ecology*, *77*(2), 513–522. https://doi.org/10.1007/s00248-018-1232-y

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*(8), 1551–1561.

Mitchell, M. W. (2011). Bias of the random forest out-of-bad (OOB) error for certain input parameters. *Open Journal of Statistics*, *1*(3), 8072. https://doi.org/10.4236/ojs.2011.13024

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nguyen, P. N., & Rehan, S. M. (2022a). Developmental microbiome of the small carpenter bee, *Ceratina calcarata*. *Environmental DNA*, *4*, 808–819. https://doi.org/10.1002/edn3.291

Nguyen, P. N., & Rehan, S. M. (2022b). The effects of urban land use gradients on wild bee microbiomes. *Frontiers in Microbiology*, *13*, 992660. https://doi.org/10.3389/fmicb.2022.992660

Noël, A., Le Conte, Y., & Mondet, F. (2020). *Varroa destructor*: How does it harm *Apis mellifera* honey bees and what can be done about it? *Emerging Topics in Life Sciences*, *4*(1), 45–57. https://doi.org/10.1042/ETLS20190125

Nogrado, K., Lee, S., Chon, K., & Lee, J.-H. (2019). Effect of transient exposure to carbaryl wettable powder on the gut microbial community of honey bees. *Applied Biological Chemistry*, *62*(1), 6. https://doi.org/10.1186/s13765-019-0415-7

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: anew versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/gr.213959.116

Odanaka, K. A., & Rehan, S. M. (2019). Impact indicators: Effects of land use management on functional trait and phylogenetic diversity of wild bees. *Agriculture, Ecosystems & Environment*, *286*, 106663. https://doi.org/10.1016/j.agee.2019.106663

Oke, T. R. (1973). City size and the urban heat Island. *Atmospheric Environment*, *7*(8), 769–779. https://doi.org/10.1016/0004-6981(73)90140-6

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). Package 'vegan'. *Community Ecology Package* version 2.5-7.

Ollerton, J., Winfree, R., & Tarrant, S. (2011). How many flowering plants are pollinated by animals? *Oikos*, *120*(3), 321–3266. https://doi.org/10.1111/j.1600-0706.2010.18644.x

Otto, S. P. (2018). Adaptation, speciation and extinction in the Anthropocene. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1891), 20182047. https://doi.org/10.1098/rspb.2018.2047

Ouyang, J. Q., Isaksson, C., Schmidt, C., Hutton, P., Bonier, F., & Dominoni, D. (2018). A new framework for urban ecology: An integration off proximate and ultimate responses to anthropogenic changes. *Integrative and Comparative Biology*, *58*(5), 915–928. https://doi.org/10.1093/icb/icy110

Pereira, F. W., Carneiro, L., & Gonçalves, R. B. (2021). More losses than gains in ground-nesting bees over 60 years of urbanization. *Urban Ecosystem*, *24*, 233–242. https://doi.org/10.1007/s11252-020-01030-1

Polyanskaya, L. M., Vedina, O. T., Lysak, L. V., & Zvyagintsev, D. G. (2002). The growth-promoting effect of *Beijerinckia mobilis* and *clostridium* sp. cultures on some agricultural crops. *Microbiology*, *71*, 109–115. https://doi.org/10.1023/A:1017914803544

Potts, S. G., Imperatriz-Fonseca, V., Ngo, H. T., Aizen, M. A., Biesmeijer, J. C., Breeze, T. D., Dicks, L. V., Garibaldi, L. A., Hill, R., Settele, J., & Vanbergen, A. (2016). Safeguarding pollinators and their values to human well-being. *Nature*, *540*, 220–229. https://doi.org/10.1038/nature20588

Proesmans, W., Albrecht, M., Gajda, A., Neumann, P., Paxton, R. J., Pioz, M., Polzin, C., Schweiger, O., Settele, J., Szentgyörgyi, H., Thulke, H.-H., & Vanbergen, A. J. (2021). Pathways for novel epidemiology: Plant-pollinator-pathogen networks and global change. *Trends in Ecology & Evolution*, *26*(7), 623–636. https://doi.org/10.1016/j.tree.2021.03.006

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing https://www.R-project.org/

Raymann, K., Coon, K. L., Shaffer, Z., Salisbury, S., & Moran, N. A. (2018). Pathogenicity of *Serratia marcescens* strains in honey bees. *mBio*, *9*(5), e01649-18. https://doi.org/10.1128/mBio.01649-18

Raymann, K., & Moran, N. A. (2018). The role of the gut microbiome in health and disease of adult honey bee workers. *Current Opinion in Insect Science*, *26*, 97–104. https://doi.org/10.1016/j.cois.2018.02.012

Rehan, S. M., & Richards, M. H. (2010). Nesting biology and subsociality in *Ceratina calcarata* (Hymenoptera: Apidae). *The Canadian Entomologist*, *142*(1), 65–74. https://doi.org/10.4039/n09-056

Rho, M., Tang, H., & Ye, Y. (2010). FragGeneScan: Predicting genes in short and error-prone reads. *Nucleic Acids Research*, *38*(20), e191. https://doi.org/10.1093/nar/gkq747

Ritchie, H., & Roser, M. (2018). *Urbanization our world data*. GitHub, Inc.

Rocha-Filho, L. C., Montagnana, P. C., Araújo, T. N., Moure-Oliveira, D., Boscolo, D., & Garôfalo, C. A. (2021). Pollen analysis of cavity-nesting bees (Hymenoptera: Anthophila) and their food webs in a city. *Ecological Entomology*, *47*(2), 146–157. https://doi.org/10.1111/een.13097

RStudio Team. (2021). *RStudio: Integrated development environment for R*. RStudio, PBC http://www.rstudio.com/

Russo, L., de Keyzer, C. W., Harmon-Threatt, A. N., LeCroy, K. A., & MacIvor, J. S. (2021). The managed-to-invasive species continuum in social and solitary bees and impacts on native bee conservation. *Current Opinion in Insect Science*, *46*, 43–49. https://doi.org/10.1016/j.cois.2021.01.001

Sánchez-Echeverría, K., Castellanos, I., Mendoza-Cuenca, L., Zuria, I., & Sánchez-Rojas, G. (2019). Reduced thermal variability in cities and its impact on honey bee thermal tolerance. *PeerJ*, *7*, e7060. https://doi.org/10.7717/peerj.7060

Santamaria, J., Villalobos, E. M., Brettell, L. E., Nikaido, S., Graham, J. R., & Martin, S. (2018). Evidence of *Varroa*-mediated deformed wing virus spillover in Hawaii. *Journal of Invertebrate Pathology*, *151*, 126–130. https://doi.org/10.1016/j.jip.2017.11.008

Schoonvaere, K., De Smet, L., Smagghe, G., Vierstraete, A., Braeckman, B., & de Graaf, D. C. (2016). Unbiased RNA shotgun metagenomics in social and solitary wild bees detects associations with eukaryote parasites and new viruses. *PLoS One*, *11*(12), e0168456. https://doi.org/10.1371/journal.pone.0168456

Seyedmousavi, S., Guillot, J., Arné, P., de Hoog, G. S., Mouton, J. W., Melchers, W. J. G., & Verweij, P. E. (2015). *Aspergillus* and aspergilloses in wild and domestic animals: A global health concern with parallels to human disease. *Medical Mycology*, *53*(8), 765–797. https://doi.org/10.1093/mmy/myv067

Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2017). A comparison of individual-based genetic distance metrics for landscape genetics. *Molecular Ecology Resources*, *17*, 1308–1317. https://doi.org/10.1111/1755-0998.12684

Soro, A., Quezada-Euan, J. J. G., Theodorou, P., Mortiz, R. F. A., & Paxton, R. J. (2017). The population genetics of two orchid bees suggests high dispersal, low diploid male production and only an effect of Island isolation in lowering genetic diversity. *Conservation Genetics*, *18*, 607–619. https://doi.org/10.1007/s10592-016-0912-8

Spear, D. M., Silverman, S., & Forrest, J. R. K. (2016). Asteraceae pollen provisions protect *Osmia* mason bees (Hymenoptera: Megachilidae) from brood parasitism. *The American Naturalist*, *187*(6), 797–803. https://doi.org/10.1086/686241

Spigaglia, P., Barbanti, F., Marocchi, F., Mastroleo, M., Baretta, M., Ferrante, P., Caboni, E., Lucioli, S., & Scortichini, M. (2020). *Clostridium bifermentans* and *C. subterminale* are associated with kiwifruit vine decline, known as "moria", in Italy. *Plant Pathology*, *69*(4), 765–774. https://doi.org/10.1111/ppa.13161

Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics*, *123*(3), 597–601. https://doi.org/10.1093/genetics/123.3.597

Tarpy, D. R., Mattila, H. R., & Newton, I. L. G. (2015). Development of the honey bee gut microbiome throughout the queen-rearing process. *Applied and Environmental Microbiology*, *81*(9), 3182–3191. https://doi.org/10.1128/AEM.00307-15

Tehel, A., Brown, M. J. F., & Paxton, R. J. (2016). Impact of managed honey bee viruses on wild bees. *Current Opinion in Virology*, *19*, 16–22. https://doi.org/10.1016/j.coviro.2016.06.006

Theodorou, P., Albig, K., Radzevičiūtė, R., Settele, J., Schweiger, O., Murray, T. E., & Paxton, R. J. (2016). The structure of flower visitor networks in relation to pollination across an agricultural to urban gradient. *Functional Ecology*, *31*(4), 838–847. https://doi.org/10.1111/1365-2435.12803

Theodorou, P., Baltz, L. M., Paxton, R. J., & Soro, A. (2021). Urbanization is associated with shifts in bumblebee body size, with cascading effects on pollination. *Evolutionary Applications*, *14*(1), 53–68. https://doi.org/10.1111/eva.13087

Theodorou, P., Radzevičiūtė, R., Kahnt, B., Soro, A., Grosse, I., & Paxton, R. J. (2018). Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (*Bombus lapidaries* L.). *Proceedings of the Royal Society B: Biological Sciences*, *285*(1877), 20172806. https://doi.org/10.1098/rspb.2017.2806

Theodorou, P., Radzevičiūtė, R., Settele, J., Schweigher, O., Murray, T. E., & Paxton, R. J. (2016). Pollination services enhanced with urbanization despite increasing pollinator parasitism. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1833), 20160561. https://doi.org/10.1098/rspb.2016.0561

Traynor, K. S., Mondet, F., de Miranda, J. R., Techer, M., Kowallik, V., Oddie, M. A. Y., Chantawannakul, P., & McAfee, A. (2020). *Varroa destructor*: A complex parasite, crippling honey bees worldwide. *Trends in Parasitology*, *36*(7), 592–606. https://doi.org/10.1016/j.pt.2020.04.004

van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, *76*(13), 1–21. https://doi.org/10.18637/jss.v076.i13

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth ed.). Springer.

Voulgari-Kokota, A., McFrederick, Q. S., Steffan-Dewenter, I., & Keller, A. (2019). Drivers, diversity, and functions of the solitary-bee microbiota. *Trends in Microbiology*, *27*(12), 1034–1044. https://doi.org/10.1016/j.tim.2019.07.011

Walterson, A. M., & Stavrinides, J. (2015). *Pantoea*: Insights into a highly versatile and diverse genus within the Enterobacteriaceae. *FEMS Microbiology Reviews*, *39*(6), 968–984. https://doi.org/10.1093/femsre/fuv027

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569. https://doi.org/10.1038/ng.608

Yang, P., Ren, G., Yan, P., & Deng, J. (2021). Urbanization reduces frequency of light rain: An example from Beijing city. *Theoretical and Applied Climatology*, *145*, 763–774. https://doi.org/10.1007/s00704-021-03655-4

Yoder, J. A., Jajack, A. J., Rosselot, A. E., Smith, T. J., Yerke, M. C., & Sammataro, D. (2013). Fungicide contamination reduces beneficial fungi in bee bread based on an area-wide field study in honey bee, *Apis mellifera*, colonies. *Journal of Toxicology and Environmental Health, Part A*, *76*, 587–600. https://doi.org/10.1080/15287394.2013.798846

Youngsteadt, E., Appler, R. H., López-Uribe, M. M., Tarpy, D. R., & Frank, S. D. (2015). Urbanization increases pathogen pressure on feral and managed honey bees. *PLoS One*, *10*(11), e0142031. https://doi.org/10.1371/journal.pone.0142031

Zheng, H., Perreau, J., Powell, J. E., Han, B., Zhang, Z., Kwong, W. K., Tringe, S. G., & Moran, N. A. (2019). Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(51), 25909–25916. https://doi.org/10.1073/pnas.1916224116

Zurbuchen, A., Cheesman, S., Klaiber, J., Müller, A., Hein, S., & Dorn, S. (2010). Long foraging distances impose high costs on offspring production in solitary bees. *Journal of Animal Ecology*, *79*(3), 674–681. https://doi.org/10.1111/j.1365-2656.2010.01675.x

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: Chau, K. D., Samad-zada, F., Kelemen, E. P., & Rehan, S. M. (2023). Integrative population genetics and metagenomics reveals urbanization increases pathogen loads and decreases connectivity in a wild bee. *Global Change Biology*, *29*, 4193–4211. https://doi.org/10.1111/gcb.16757