Impact of historical and current features of the cityscape on the genetic structure of the house mouse (*Mus musculus domesticus*) in Dakar (Senegal, West Africa).

Claire Stragier[1], Sylvain Piry[2], Anne Loiseau[2], Mamadou Kane[1], Aliou Sow[1], Youssoupha Niang[1], Mamoudou Diallo[1], Arame Ndiaye[1], Philippe Gauthier[3], Marion Borderon[4], Laurent Granjon[3], Carine Brouat[3,6*], Karine Berthier[5,6*]

[1] BIOPASS (IRD-CBGP, ISRA, UCAD), Campus de Bel-Air, BP 1386, CP 18524 Dakar, Senegal.

[2] CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

[3] CBGP, IRD, CIRAD, INRA, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

[4] Department of Geography and Regional Research, University of Vienna, Austria

[5] Pathologie Végétale, INRA, 84140 Montfavet, France

[6] These two authors contributed equally to this work.

***Corresponding author:** Carine Brouat, CBGP, Montferrier sur Lez, France. Tel: +33 4 99 62 33 09; Fax: +33 4 99 62 33 45; E-mail: carine.brouat@ird.fr

Running title: House mouse population genetics in Dakar

Abstract (184 words)

Population genetics of human commensal species within cities may reflect current features of the cityscape, including socio-economic aspects, as well as urbanization history. Understanding genetic structure of archetypal commensals such as rodent pests may provide useful insights to manage their populations. In this study, we addressed these issues by focusing on the house mouse *Mus musculus domesticus* in Dakar, Senegal. The house mouse is a major synanthropic rodent who may have been introduced as soon as Dakar was settled by Europeans at the middle of the nineteenth century. We examine genetic variation at one mitochondrial locus and 15 nuclear microsatellite markers, in mice from 14 sampling sites representing different stages of urbanization history and different socio-economic environments in Dakar. We further investigate the influence of the historical and current features of the cityscape of Dakar on genetic structure of the house mouse, using a landscape genetic approach. Genetic patterns were consistent with an introduction of the house mouse at colonial times. Spatial analyses showed that differentiation reflects both urbanization history and current urban features, which may have practical implications in terms of rodent management.

Keywords: *Mus musculus domesticus*, invasion biology, socio-economic urban habitat, cityscape, population genetics, landscape genetics, microsatellite, D-loop

INTRODUCTION

Urbanization is a major environmental change and represents the land use that is the fastest growing (Bloom *et al.* 2008), especially in developing countries (Cohen 2006, Chen *et al.* 2014). Urbanization is considered as an extreme type of habitat fragmentation that can lead to spatial isolation of small subpopulations or even to local extinction (McKinney 2006, Munshi-South *et al.* 2016). As a consequence, genetic drift is expected to be enhanced for numerous species, leading to a decrease of genetic diversity within local subpopulations and to an increase of genetic differentiation between them (Johnson & Munshi-South 2017). For commensal species relying on human habitats and resources, the situation can be quite different as they are not limited to remnant patches of native habitat and can easily inhabit, and disperse through, a diversity of urban infrastructures (Combs *et al.* 2018b, Combs *et al.* 2018a).

Invasive rats and mice are archetypal commensals and a major concern for public health, food security and biodiversity all around the world (Singleton *et al.* 2003, Singleton *et al.* 2015; Angel *et al.* 2009, Meerburg *et al.* 2009, Parsons *et al.* 2017). The cost of rodent control actions weighs heavily on city's budget, and classical control protocols, relying mostly on poisoning, have proven to be quite inefficient in reducing urban rodent densities in the long term (Parsons *et al.* 2017, Richardson *et al.* 2017). It is now well recognized that the development of more efficient management strategies requires a better understanding of urban rodent ecology (Singleton *et al.* 1999, Makundi & Massawe 2011, Feng & Himsworth 2014, Parsons *et al.* 2017, Combs *et al.* 2018b, Byers *et al.* 2019). Understanding spatio-temporal variation in gene flow is clearly critical to help in designing appropriate control strategies (Gardner-Santana *et al.* 2009, Kajdacsi *et al.* 2013, Johnson & Munshi-South 2017, Parsons *et al.* 2017, Richardson *et al.* 2017, Combs *et al.* 2018b, Combs *et al.* 2018a). For instance, classic implementation of rodent control at the scale of a single building or block is likely to fail if genetic patterns suggest a high probability of recolonization from surrounding groups. Conversely, identifying clear genetic clusters can allow defining meaningful spatial units for control actions (Kajdacsi *et al.* 2013, Richardson *et al.* 2017, Combs *et al.* 2018b). As anticoagulant resistance can rapidly reach high frequency within local rodent

populations (Berny *et al.* 2018), knowledge on spatial patterns of gene flow can also help in predicting the risk at which resistance alleles can spread across a cityscape (Haniza *et al.* 2015).

So far, population genetic studies of urban rodents have mostly focused on the brown rat (*Rattus norvegicus*). Several of these studies showed that gene flow occurs at a sufficient rate to limit genetic drift and that spatial genetic variation tends to be organized under isolation-by-distance (IBD) resulting from both the spatially restricted dispersal and social behavior of this species (Gardner-Santana *et al.* 2009, Combs *et al.* 2018b, Combs *et al.* 2018a). Beyond this general trend, spatially explicit analyses have also revealed sharp genetic discontinuities at relatively small spatial scales (Richardson *et al.* 2017, Combs *et al.* 2018a). Such genetic breaks have been attributed to physical barrier to dispersal (waterway in Baltimore, USA - Gardner-Santana *et al.* 2009; road and topography in Pau de Lima, Salvador, Brazil - Kajdacsi *et al.* 2013, Richardson *et al.* 2017) or spatial heterogeneity in habitat quality that impact local population size (resource availability varying with socioeconomic features of urban infrastructures, Manhattan, USA - Combs *et al.* 2018a). Several studies have advocated for the need to integrate socioeconomic features of the urban environment in population dynamic and genetic studies of commensal rodents in urban landscape (Masi *et al.* 2010, van Adrichem *et al.* 2013, Lucaccioni *et al.* 2016, Lambert *et al.* 2017, Santos *et al.* 2017, Combs *et al.* 2018b, Byers *et al.* 2019). In addition to current features, spatial genetic patterns may also reflect past demographic events related to cityscape evolution during the historical timeframe of urbanization (Harris *et al.* 2016, Lourenço *et al.* 2017). This can be especially relevant for non-native invasive commensal rodents that were introduced in numerous primary European settlements which grew into cities during the period of colonial expansion (Aplin *et al.* 2011, Jones *et al.* 2013, Puckett *et al.* 2016).

In this study, we investigate the influence of the historical and current features of the cityscape of Dakar, Senegal, on the spatial pattern of genetic variation of the house mouse, *Mus musculus domesticus*, a major invasive rodent pest recorded in cities of all continents. This subspecies from the *Mus musculus* complex originates from South-West Asia (Suzuki *et al.* 2013), and became commensal during the initial settlements of humans in the Middle East at circa 10 000 BC (Cucchi *et al.* 2012). From there, it has expanded its distribution to reach a cosmopolitan range (Bonhomme & Searle 2012). Several

phylogeographic studies have shown that its colonisation history is strongly related to human migration patterns at different timescales (Jones *et al.* 2013). The house mouse may have been introduced in Dakar as soon as it was settled by Europeans at the middle of the nineteenth century (Lippens *et al.* 2017). Dakar has been continuously growing since European settled in the southern end of the Cap-Vert peninsula. From the late nineteenth century to the First World War, urban development was driven by the will to improve commercial links between France and its West African colonies (Harris 2011). This led to the early creation of a commercial area around the port that further expanded along the railway that connected Dakar to Saint-Louis (the capital city of former French West Africa in the North-West of Senegal) in 1885. The spatial expansion of the city has accelerated after World War II and, in absence of sound urban planning, has resulted in a highly fragmented urban matrix in terms of economic activities and housing, e.g. residential, planned, spontaneous legal and spontaneous illegal housing (Diallo *et al.* 2012, Borderon *et al.* 2014). Today, the city is the political, cultural and economic capital of Senegal (Rouhana *et al.* 2015). It is also one of the leading industrial, financial and service center of West Africa, and a major seaport of the western African coast.

We used two different types of genetic markers to characterize the genetic variation of the house mouse in Dakar: sequences from the mitochondrial DNA control region (D-loop), and 15 nuclear microsatellites. Data on the D-loop, which is the only molecular marker for which substantial data are available over the entire distribution of the taxon (Lippens *et al.* 2017), were used to further investigate the geographic origin of house mouse populations introduced in Dakar. Microsatellite data, which allow characterizing fine-scale population genetic structure, were used to determine whether spatial genetic structure is rather consistent with an isolation-by-distance process or with well-defined genetic units (in the sense of model-based genetic clustering). We further investigated the effects of the spatio-temporal process of urbanization along with the current landscape and socioeconomics features of Dakar's cityscape using the MAPI method (Piry *et al.* 2016) and a conditional autoregressive model.

MATERIALS AND METHODS

Study area

Dakar is located on the Cap-Vert Peninsula on the Atlantic coast (14°41′34″N 17°26′48″W) (Fig. 1). It has a tropical climate with a rainy season from June to October, average annual rainfall of ca. 500 mm, and average temperatures between 17-31°C.

The development of the city has gone through several stages (Supplementary information, Fig. S1). The first colonial settlement was founded in 1621 on Gorée Island (Sinou 1993). At this period, the Cap-Vert Peninsula was settled by an ethnic group called Lebou, distributed in small villages on the coast. The south cape of the Peninsula (actual Plateau district) was colonized by Europeans since 1857, but the urban settlement actually began with the achievement of a railway from Saint-Louis to Dakar in 1885 (Sinou 1993). At the independence in 1960, the city covered only the southern half of the Cap-Vert peninsula (Vernière 1973, Salem 1981). It then extended rapidly to include the ancient fishermen villages of Ouakam, Yoff and Ngor located in the North and to eventually cover the whole peninsula (except for the airport area). This urban expansion was related to a considerable demographic growth, like in many other cities of the developing countries (Cohen 2006). For instance, the population of Dakar city increased from 530 000 inhabitants in 1976 (Pélissier 1983) to 1,15 million in 2013 (ANSD 2016).

Spatio-temporal pattern of urbanization

The spatio-temporal dynamics of urbanization in the Cap-Vert peninsula was assessed using: (i) historical data to identify the localization of built-up areas at twelve time points over a 156 year-period, from 1862 to present (Supplementary information, Fig. S1); and (ii) results of a recent classification of Dakar urban habitats based on satellite images and census data (Borderon *et al.* 2014).

Eleven historical maps and one recent map were retrieved from various web sources or from paper maps provided by the Senegalese national geographical institute (see details in Supplementary Material, Appendix 1). All images were georeferenced as “rasters” using the QGis software (QGIS Development Team 2018). As coordinates were not always available or reliable, easily identifiable features such as coastlines, major road crossings and outstanding buildings were used as control points for the georeferencing process. Once georeferenced, built-up areas from each map were hand digitized as

polygons in QGis, using WGS84 (EPSG:4326) latitude/longitude coordinates, and exported as shapefiles to a PostgreSQL/PostGIS database using the R package “sf” (Pebesma 2018).

To describe the current features of urban habitats within the peninsula, we used the results of the supervised classification of land cover realized by Borderon et al. (2014). For this classification, the authors used SPOT 5 images (recorded in September 2007), which were processed in order to obtain a final spatial resolution of 2.5m. Each pixel was automatically attributed to 16 classes of land cover defined using training units based on field knowledge. Among these 16 classes, six represented the socio-economic profiles of the urban habitats (Supplementary information, Table S1, Fig. S2) defined by correlating the habitat typology proposed by the United Nations (Un-Habitat 2008) and the results of an ascending hierarchical classification applied on census data on population wealth (see Borderon et al. 2014 for details). The other classes included five types of bare ground, one class of strict vegetation, two different classes of vegetation mixed with bare ground and two types of water cover. In our study area, the five classes of bare ground and the two classes of water cover corresponded to very local or under-represented land covers with no obvious ecological meaning for the house mouse (i.e. beaches, linear pavement, extraction sites, small water body…), and were thus discarded. For further analyses, we only considered the six classes of urban habitats based on socio-economic profiles, and merged the three classes including vegetation (strict or mixed with bare ground) into a single class called “Vegetation”.

Sampling

Within Dakar, we selected 14 sampling sites separated from each other by a median distance of 5850 meters (min: 660 meters; max: 15690) (Table 1; Fig. 1). They were chosen to represent different stages of the urban history and different habitats (Supplementary information, Table S1, Fig. S1, Fig. S2). IDG, PRB and PFD were located in the first urban areas developed by Europeans in Gorée Island (IDG) or in the extreme South of the Cap-Vert peninsula (PRB and PFD). POR corresponds to the commercial harbor of Dakar. GTP corresponds to a first extension of the city that was developed in 1914 for the indigenous population during epidemic plague (M’bokolo 1982). GDK, CBN, and PDO are within districts that were built from 1940. YOF, OKM, NGR and HPE are inside traditional fisherman villages of Lebou that have

been successively absorbed by the urban development after 1960 (Supplementary information, Fig. S1). PAH, a public and zoological park, is the only large green space within this urban context.

Fieldwork was carried out under the framework agreements established between the Institut de Recherche pour le Développement and the Republic of Senegal, as well as with the Senegalese Head Office of Waters and Forests, and the Ministry of Health and of the Social Action. Handling procedures were performed under our lab agreement for experiments on wild animals (no. D-34-169-1), and follow the official guidelines of the American Society of Mammalogists (Sikes *et al.* 2011). Trapping campaigns within district were systematically performed with prior explicit agreement from relevant local authorities.

Sampling was performed in 2016, according to a standardized protocol already described (Dalecky *et al.* 2015). Each sampling site was the focus of one live trapping session of 2 to 5 consecutive days, in order to reach the target sample size of 20 captured house mice on a median surface of 0.04 km² (min: 0.01, max: 1.51; the maximum value concerned POR). In general, two traps (one wire-meshed trap and one Sherman trap) were set per room or courtyard in buildings corresponding to dwelling houses, boutiques, workshops, offices or warehouses, and whose location was precisely recorded with a GPS device. In POR, trapping was done inside and around 19 offices and warehouses (between 5 and 20 pairs of traps per warehouse). House mice were brought back to our laboratory, euthanized by cervical dislocation, weighted, sexed and measured the same day. A piece of tissue (hind foot) was stored in 95% ethanol for molecular analyses.

Laboratory analyses

For the 529 mice sampled, genomic DNA was extracted from tissues using the DNeasy Blood & Tissue Kit QIAGEN according to the manufacturer's protocol.

The complete D-loop sequence was amplified for 40 mice sampled within eight of the sampled sites (from three to nine mice from distant houses per site, Table 1). We used PCR primers and conditions described in Rajabi-Maham *et al.* 2008. PCR products were sequenced in both directions by Eurofins MWG (Ebersberg, Germany).

Genotyping was performed for the 529 mice sampled using 15 nuclear microsatellites loci (D1Mit291, D2Mit456, D3Mit246, D4Mit17, D4Mit241, D6Mit373, D7Mit176, D9Mit51, D10Mit186, D11Mit236, D14Mit66, D16Mit8, D17Mit101, D18Mit8 and D19Mit30: available from the MMDBJ database: http://www.shigen.nig.ac.jp/mouse/mmdbj/top.jsp), according to Lippens *et al*. (2017).

Mitochondrial sequence analysis

All 40 D-loop sequences from Dakar were aligned with 119 sequences from Senegal, published by Lippens *et al.* (2017), and with 1958 sequences retrieved from GenBank (1672 whose references were compiled in Lippens *et al.* 2017; 168 published in other studies: see references in Supplementary information, Fig. S3). Sequences alignment was performed using the Clustal W multiple alignment in Bioedit v7.2.6 software (Hall 1999). Haplotypes were identified with DNA Sequence Polymorphism v6.10.01 software (Rozas *et al.* 2017). Those found in Dakar were classified following the haplogroup nomenclature defined by Bonhomme *et al.* 2011. A phenogram was then constructed using median joining network calculations implemented in NETWORK v5.0.0.1. (http://www.fluxus-engineering.com) on haplotypes from Senegal, in order to illustrate their relative frequency and relatedness.

Microsatellite population diversity and structure

We integrated to the microsatellite dataset 28 mice from an additional district of Dakar (DAK) that were sampled in 2014 and genotyped on the same markers in a previous study (Lippens *et al.* 2017). DAK corresponds to the district of Sacré Coeur founded in 1960 at the heart of the Cap-Vert Peninsula (Fig. 1).

Population genetic analyses were thus done on a dataset comprising 557 house mice (Table 2). Deviations from Hardy-Weinberg Equilibrium (HWE) and genotypic linkage disequilibrium (LD) were tested by locus and by sampling site using GENEPOP v4 (Rousset 2008). Corrections for multiple tests were performed using the false discovery rate (FDR) approach (Benjamini & Hochberg 1995) implemented in the Bioconductor's qvalue package (Storey 2002) of R. Heterozygote deficiencies are often found in house mouse populations and are classically attributed to their social system (Ihle et al., 2006). We examined whether null alleles or Wahlund effect may alternatively explain heterozygote deficiencies in the dataset. For this purpose, null allele frequencies were estimated using the software FREENA (Chapuis

& Estoup 2007) for each sampling site at each locus. We also calculated the kinship coefficient (p) (Loiselle *et al.* 1995) between all pairs of individuals within sampling sites, with SPAGeDI v1.4 (Hardy & Vekemans 2002). We used the genotype data from all localities as the dataset of reference for allelic frequencies.

For each sampling site, genetic diversity was estimated using the expected heterozygosity H_s (Nei, 1987), the inbreeding coefficient (F_{is}), and the allelic richness (a_r) calculated using the rarefaction procedure implemented in FSTAT v2.9.3 (Goudet 2001) for a minimum sample of 19 diploid individuals. Effective population size (N_E) was estimated for each sampling site using the linkage disequilibrium method implemented in the software NeEstimator v2 (Do *et al.* 2014) with a minimum allele frequency (i.e. *Pcrit* parameter) of 0.02. Genotypic differentiation among sampling sites was tested using Markov chain methods in Genepop (using the FDR approach to correct for multiple tests). Genetic differentiation between sampling sites was summarized by calculating pairwise F_{ST} estimates (Weir & Cockerham 1984) using FSTAT.

Allelic richness, H_S, F_{IS} and F_{ST} were compared using one sided test with Fstat (10,000 permutations) among two groups of sampling sites of the Cap-Vert peninsula. The first group corresponded to sampling sites within the first urban area developed by European settlers (PRB, PFD, GTP, POR) and the second group corresponded to ancient villages of fishermen that were more recently connected through the spatially continuous urbanization of Dakar (e.g., OKM, NGR, YOF: Supplementary information, Fig. S1). We expected to find lower genetic diversity and differentiation in the second group if it originated from the expansion of the first group.

Under a model of isolation by distance (IBD), genetic distance between populations is expected to increase with geographical distance (Rousset 1997). IBD was analysed between sampling sites by regressing pairwise estimates of F_{ST} / (1- F_{ST}) against log-transformed geographical distances. Mantel tests were performed to test the correlation between matrices of genetic differentiation and Euclidean geographical distance using GENEPOP v.4 (10 000 permutations).

We characterized the spatial genetic structure of the house mouse in Dakar using two approaches. First, we used the clustering approach implemented in STRUCTURE v2.3.4 (Pritchard *et al.* 2000), in order to estimate the number of homogeneous genetic groups (*K*) in the data set. The analyses were performed with a model including admixture and correlated allele frequencies (Falush *et al.* 2003). We performed 20 independent runs for each *K* values (from K=1 to 10). Each run included 500,000 burn-in iterations followed by 1,000,000 iterations. The number of genetic groups was inferred by the delta*K* method applied to the log probabilities of data (Evanno *et al.* 2005). We checked that a single mode was obtained in the results of the 20 runs for all K-values explored, using the Greedy Algorithm implemented in CLUMPP v.1.2.2 (Jakobsson & Rosenberg 2007). Bar plots were finally generated with DISTRUCT v.1.1 (Rosenberg 2004).

STRUCTURE results may be biased because of deviations from Hardy-Weinberg equilibrium, as well as IBD (Pritchard *et al.* 2000). We also analyzed the spatial genetic pattern of differentiation within the peninsula (i.e. Gorée Island was excluded from the dataset) using the MAPI method (Mapping Averaged Pairwise Information, implemented in the R package “mapi”) that has a low sensitivity to IBD and is not based on a predefined population genetic model (Piry *et al.* 2016). MAPI is basically a smoothing-like procedure applied on pairwise genetic measures, such as F_{ST}, which are attributed to a network of ellipses linking the sampling points. As a result, MAPI produces geographical maps representing the spatial variation in the average level of genetic differentiation between sampled localities. These genetic maps can be superimposed on landscape layers to further analyze the relationship between genetic structure and landscape features (Piry *et al.* 2018; Dellicour *et al.* 2019). As the spatial distribution of sampled mice was rather clustered, roughly representing 13 sampling sites, MAPI was run at the population- rather than individual-level using F_{ST} values computed between all pairs of localities to build the spatial genetic network of ellipses required by the analysis. To account for the dispersion of individuals within each of the 13 sampled sites, an error circle radius of 300 meters was applied to the spatial coordinates of all sites, except for the harbor POR (radius = 1000 m) and the Hann park ‘PAH’ (radius = 500 m). A grid of 243 hexagonal cells with a half-width of 310 m was superimposed on the study area. This resolution was determined by setting the ‘beta’ parameter to 0.25 in the MAPI analysis in

respect of the Nyquist-Shannon sampling theorem under a situation of random sampling (see Piry *et al.* 2016 for further details). Each cell received the weighted arithmetic mean of the F_{ST} values attributed to the ellipses intercepting its geographical extent. This mean was weighted using the inverse of the ellipse areas to limit long distance effects (i.e., IBD). The eccentricity value of the ellipses, which controls the smoothing intensity, was set to the default value of 0.975 (see Piry *et al.* 2016 for further details on the effects of this parameter). We used the permutation procedure (10,000 permutations), along with the FDR approach as described in Piry et al. (2016), to identify major areas of genetic continuity and discontinuity, e.g., areas exhibiting, lower or higher levels of genetic differentiation, respectively, than expected randomly.

To analyze the relationship between the spatial genetic structure of house mouse populations and the urbanization process of Dakar, we superimposed the MAPI grid on the layers describing the historical and current cityscape. First, we determined the localization of the first European settlement (spatial coordinates: 14.67210°N 17.429235°W) from the city map of 1850 (Fig. 2). We also followed the approach described in Brouat *et al.* 2014 to identify the geographical location of the initial introduction site by regressing the expected heterozygosity (H_S), computed for each sampling site, on the geographical Euclidean distance separating this site to the center of each cell of the MAPI grid. The cells with negative regression slopes and the highest adjusted r-squared (r^2) indicated the most likely site from which the house mouse has invaded the peninsula (Ramachandran *et al.* 2005). To assess the significance of the correlation, the spatial location of the sampling sites were permuted (20,000 random permutations). Only the cells with a r^2 higher than the 95th quantile of the permuted distribution were retained and mapped. Second, we used the time series of 12 maps of Dakar (Supplementary information, Fig. S1) to attribute a first date of urbanization to each cell of the MAPI grid. This variable called “built-up” was determined by considering that built-up polygons must cover at least 5% of a cell surface for the latter to be considered as urbanized. To allow further analyses, we arbitrary attributed the date of 2025 to the seven cells that could still not be identified as urbanized from the 2018 map (Fig. 2). Once a cell was classified as urbanized we subsequently determined at which date it was spatially connected to the first European settlement through continuous urban areas (i.e. pathway across built-up cells). This variable called

“connection” provides temporal information on the spatial expansion of the city from the first European settlement. The relationships between the historical variables (“built-up” and “connection”) and the averaged value of F_{ST} obtained within each cell of the MAPI grid was analyzed using Spearman rank correlations.

Within each cell of the MAPI grid, we also computed the proportion of each of the seven land cover classes (i.e. 6 classes of socio-economic urban habitat and one class of vegetation) retrieved form the classification analysis of Borderon *et al.* (2014) (Fig. 2; Supplementary information, Table S1, Fig. S2). We finally used a conditional autoregressive model (CAR) to analyze the relationships between MAPI results (i.e. averaged F_{ST} value within each cell) and the frequency of each of the seven land cover classes as well as the standardized time period between the year during which most of the sites have been sampled (2016) and both, the year at which a cell is considered as urbanized (variable “built-up”) and the year at which a cell is connected to the first European Settlement (variable “connection”). To account for spatial autocorrelation, the error term of the model was assumed to include two components: an unstructured component (i.e. independently distributed) and a structured component (i.e. spatially correlated). The unstructured heterogeneity component was assumed to be centered and normally distributed while the structured heterogeneity component was assumed to have a conditional intrinsic Gaussian autoregressive (CAR) distribution (Besag *et al.* 1991) with an inverse-distance neighborhood weight matrix. The model was fitted in a Bayesian framework with an INLA approach (Blangiardo & Cameletti 2015) using the R-INLA package (Martins *et al.* 2013). Bayesian inference resulted in posterior densities for all predictors from which significance was assessed with a threshold of 0.05. To evaluate the fit of the model to the data we computed Conditional Predictive Ordinate (CPO) and Probability Integral Transformation (PIT) values (see Held *et al.* 2010). We checked for the occurrence of extremely small values of CPO indicating unlikely observations under the fitted model. We used a Kolmogorov-Smirnov test to confirm that PIT values did not significantly depart from a uniform distribution as expected if model assumptions are correct. As numerical problem can occur when computing CPO and PIT, the “inla” function of the R-INLA package provides a measure of the reliability of each value which is scaled between 0 and 1 (values far from 0 indicate that for a given observation, the corresponding predictive

measure is not reliable). Finally, we also checked that model residuals were not overly structured in space.

RESULTS

The house mouse was the most sampled small mammal in all sampling sites (data not shown). Sample sizes for genetic analyses ranged from 19 to 72 mice per sampling site (Table 1).

Mitochondrial sequence analysis

In the 40 D-loop sequences obtained from our sampling sites and the 5 sequences previously obtained from an additional district (DAK) of Dakar, there were 9 variable sites defining 5 haplotypes (Table 1; mean haplotype diversity: 0.264 ± 0.313; mean nucleotide diversity: 0.003 ± 0.005; Supplementary information, Fig. S3). Two haplotypes corresponded to haplotypes H1 (80% of the individuals) and H2 (9% of individuals) previously found by Lippens et al. (2017) to be major D-loop haplotypes in Senegal: they belonged to the haplogroups HG11 and HG4, respectively, in the nomenclature defined by Bonhomme et al. (2011). H1 was found in all sampling sites except PFD, and H2 in three sampling sites (DAK, GTP and PFD). The three other haplotypes (H16, H17 and H18) were not previously found in Senegal. H17 (2.2% of individuals, only found in GTP) and H18 (2.2% of individuals; only found in IDG) were separated from H1 by only a few mutational steps (3 and 1, respectively) and belonged also to HG11. H16 (6.7% of the individuals) was found in POR and PFD and belonged to haplogroup HG1 that were already found in one individual of Saint-Louis, North Senegal (Lippens *et al*. 2017). Haplotype diversity was greater in sampling sites surrounding the first European settlements (GTP, PFD, POR, IDG), than in old fisherman villages (OKM, HPE, YOF, NGR: only H1) (Table 1; Supplementary Information, Fig. S3).

Microsatellite analyses

Genetic analyses were conducted on the dataset regrouping the 529 house mice sampled for this study and the 28 mice previously sampled in DAK (Supplementary information, Table S2). Linkage disequilibrium

was significant for 10 of the 1470 tests performed (0.7%) and the significant values involved different pairs of loci. The 15 loci were thus considered to be genetically independent. None of the loci were at Hardy-Weinberg equilibrium after FDR correction, and positive F_{IS} values were obtained at all sites (Supplementary information, Table S2). The locus D4m17 displayed 11 null genotypes, and its mean F_{IS} (0.31) was significantly higher than those of the other loci (mean F_{IS}= 0.22, 95%CI= [0.20;0.24]). This suggested the occurrence of null alleles at this locus. The mean null allele frequency estimated with FreeNA for D4m17 was however sufficiently low (0.12) not to affect significantly genetic diversity and differentiation estimates (Chapuis & Estoup 2007, Chapuis *et al.* 2008), and hence this locus was conserved in the analyses. Null alleles were unlikely to explain heterozygote deficiencies at other loci, since only a small number of null genotypes were observed (0-2 per locus). Within sampling sites, the median kinship Loiselle coefficient *p* calculated between all pairs of individuals ranged from -0.021 to 0.201, the higher value being found for Gorée Island (Table 1). Distributions of *p* within sites of the peninsula were largely unimodal (Supplementary information, Fig. S4) and globally zero-centred. This suggests that the observed deviations from Hardy-Weinberg equilibrium result from the presence, within each site, of different social breeding groups from which only a few individuals have been sampled, rather than from a strong family or Wahlund effect due to the oversampling of only one or a few of these groups (Dobson 2007, Berthier *et al.* 2016).

Allelic richness (a_r) ranged from 3.16 to 5.71 (mean 4.88 ± 0.68), H_s from 0.44 to 0.68 (mean 0.61 ± 0.06), and F_{IS} from 0.11 to 0.31 (mean 0.21 ± 0.01) per sampling site (Table 1). Allelic richness was significantly lower (P = 0.045) in sampling sites corresponding to ancient villages (OKM, NGR, YOF) than in sampling sites around the first European settlement (PRB, PFD, GTP, POR).

Genotypic differentiation was significant between sampling sites ($\chi^2_{(30)}$ = 478.27; P = 0.0001). Pairwise F_{ST} values ranged from 0.01 to 0.24 with a global mean F_{ST} value of 0.09 (95% CI= [0.08-0.10]) (Supplementary information, Table S3). As expected, substantial genetic structure was observed between Gorée Island (IDG) and all other sites sampled in the Cap-Vert peninsula (mean F_{ST} = 0.18 ± 0.03). IBD

was significant between sampling sites with IDG (Mantel test: P = 0.007; slope b = 0.046, 95% CI= [0.034-0.64]) and without IDG (Mantel test: P = 0.024; slope b = 0.018, 95% CI= [0.011-0.028]).

Spatial genetic structure was further characterized with STRUCTURE. The highest delta$K$ value was obtained for K = 2 (deltaK = 1172 at K = 2, and below 3.8 otherwise). At K=2, IDG was assigned to a first group, and all individuals from sampling sites of the Cap-Vert peninsula were largely assigned to a second group (Fig. 1A). At K = 3, the two genetic groups that emerged in the Cap-Vert peninsula were not stable among runs, and many individuals have mixed membership in both groups.

In the Cap-Vert Peninsula, results of the MAPI analysis clearly showed a longitudinal gradient from lowest averaged F_{ST} values in the eastern part of the study area to highest values in the western part (Fig. 1B). A significant area of genetic discontinuity encompassed the northern part of the Ouakam district, between the two main runways of the Yoff Airport, while a significant area of high genetic continuity (i.e. low averaged F_{ST} values) was found in the harbor area close to the first European settlement. MAPI averaged F_{ST} values were significantly and positively correlated to both the “built-up” (p = 1.707e-07) and “connection” (p < 2.2e-16) variables, suggesting that the older the urbanization or connection to the first European settlement, the lower the genetic differentiation (Supplementary information, Fig. S5). Note that the Spearman coefficient was two times greater when considering the “connection” (Rho = 0.65) rather than the “build-up” variable (Rho = 0.33).

Results of the Bayesian conditional autoregressive model (Fig. 3) revealed significant but opposite effect of the historical variables, considered here as the time period elapsed since the occurrence of built-up areas and since the connection of local built-up areas to the first European settlement. The former had a positive effect (i.e. the longer the time period since the construction of identifiable buildings, the higher the average level of genetic differentiation) while the later had a negative effect (i.e. the longer the time period since the connection of the built-up areas to the first European settlement, the lower the average level of genetic differentiation). Among the land cover classes describing the current cityscape, the classes “Vegetation” and “Urban 4” had a clear significant positive effect on the average level of differentiation (i.e. the higher the frequency of these classes, the higher the average level of genetic

differentiation) while the class “Urban 5” and more marginally the class “Urban 1” had a significant negative effect (i.e. the higher the frequency of these classes, the lower the average level of genetic differentiation). The reliability measure for the CPO and PIT values was equal to 0 in all cases, indicating that CPO and PIT could be used to assess the calibration of the model. There was no evidence for a lack of fit of the model from the CPO values as they were all superior to 1. Moreover, the distribution of the PIT values did not significantly depart from a uniform distribution (Kolgomorov-Smirnov test p-value =0.522), which confirmed that the model was well calibrated (Supplementary information, Fig. S6). Finally, model residuals were not overly structured in space (Supplementary information, Fig. S7).

DISCUSSION

In this study, we aimed to investigate the population genetic structure of *M. m. domesticus* in the city of Dakar, with both mitochondrial sequences and microsatellite markers. By linking genetic patterns with spatial and temporal data describing the urbanization process, we wanted to evaluate whether the genetic structure of this synanthropic species was reflecting both the historical dynamics of urbanization and the heterogeneity of the current cityscape.

The date of introduction of *M. m. domesticus* in Dakar is not known. The first records of the house mouse precisely in Dakar were from the early 20th century (see references in Dalecky et al. 2015). However, historical inventories of rodent fauna reported large populations of mice in colonial coastal cities of Senegal (without mention of specific localities) from the middle of the 19th century (Temminck 1853), which corresponds to the foundation of the first European settlement in the Cap-Vert Peninsula. Genetic data of house mice from Dakar support the idea of an introduction of the species during colonial times. Indeed, all the mitochondrial haplogroups found in Dakar are typical from Western Europe (Supplementary information, Table S4; Bonhomme *et al.* 2011) and the highest diversity levels for both mitochondrial markers and microsatellites were found in sampling sites at the south of the Cap-Vert peninsula (Table 1), which correspond to the first continental European settlement of the area. Before Europeans settled in the Cap-Vert peninsula, a first European settlement was founded by Dutch in Gorée

Island during the 17th century, and alternatively occupied by French and British during the 18th century (Sinou 1993). The house mouse could thus have been first introduced in Gorée before having colonized the Cap-Vert peninsula. Mitochondrial and microsatellite genetic diversities at Gorée that are embedded in those found in the Cap-Vert peninsula, would rather suggest the inverse hypothesis. However, the confounding effect of genetic drift related to insularity, which also explains the differentiation found in microsatellite data between mice from IDG and from the Cap-Vert peninsula, prevents to conclude about the direction of colonisation between the insular and continental parts of the colonial city.

House mice could have been introduced in Dakar (Gorée or the Cap-Vert peninsula) either directly from Western Europe through maritime colonial trade or from other locations of Senegal already invaded from Europe. In line of this latter hypothesis, a previous analysis of microsatellites indicated the city of Saint-Louis as the putative source of all Senegalese mice (Lippens *et al.* 2017). Saint-Louis was the first colonial port to be developed in Senegal (Sinou 1993). This city was linked to Dakar by railway from 1885 to 1933 (Tixier 1965, Bonnardel 1992), and the terminal railway station was located in the vicinity of the first European settlement in the Cap-Vert peninsula. Interestingly, the major mitochondrial haplotypes found in Dakar (H1 and H2) had been reported at relatively high frequencies (> 20%) in Great Britain, which controlled Saint-Louis for 80 years during the eighteen century (Sinou 1993) (Supplementary information, Table S4).

Following introduction, spatial expansion may be characterized by sequential founder events, leading to spatial increase of genetic structure and decrease of genetic diversity along colonization axes (Clegg *et al.* 2002). Such a pattern was shown for the house mouse at the scale of its spatial distribution in Senegal (Lippens *et al.* 2017). Although genetic structure at microsatellite markers was lower in Dakar than across Senegal, several aspects suggest a pattern of spatial expansion that is consistent with spatial urban growth from the first European settlement. First, the regression analysis based on the expected heterozygosity also pointed out the harbor area, which is contiguous to the first European settlement, as the most likely geographical origin of the house mouse expansion in the Cap-Vert peninsula. Second, when excluding Gorée Island, genetic diversity levels were the lowest in the sampling sites OKM, NGR, YOF and HPE,

which correspond to the ancient Lebou villages that have been the most recently connected to the rest of the city (Table 1). Third, results of the Bayesian CAR model suggest that the longer the time period elapsed since the connection of a built-up area to the first European settlement, the lower the average level of genetic differentiation. Interestingly, an opposite effect was found for the time-period since the first occurrence of built-up areas; the longer the time-period, the higher the average level of genetic differentiation. This discrepancy between the two historical variables describing the urbanization dynamics is due to the oldest Lebou villages located in the north part of the peninsula (especially OKM and NGR), which exhibit the highest F_{ST} values with other sites (Fig. 1; Supplementary information Table S3) and were identified as urbanized areas from the oldest maps (e.g., 1862) while being also part of the last areas connected to the old city (Fig. 2; Supplementary information Fig. S5). Altogether these results suggest that even if some areas have been inhabited by humans since a long time, their colonization by the house mouse may have occurred step-by-step with the continuous spatial expansion of the city, rather than through long-distance migration.

While the spatial genetic pattern of house mouse populations in Dakar reflects the historical dynamics of urbanization, it is also shaped by the heterogeneity of the current cityscape as suggested by the results of the Bayesian CAR model. Indeed, variables that were not representing housing *per se*, but industrial areas ("Urban 5") or vegetation had a significant effect on the average level of differentiation computed using MAPI. Industrial areas had a negative effect suggesting that gene flow is enhanced in this type of habitat, which is likely to result from high densities of house mice. Indeed, these areas include numerous warehouses to stock different food products such as groundnuts, rice, sugar and flour (Ba *et al.* 2013) that can provide resources allowing sustaining large populations. Vegetated areas were shown to be associated with higher levels of genetic differentiation, suggesting that they may act as a barrier to dispersal for the house mouse and/or as an unfavorable habitat for this synanthropic species, in such a way that local densities are very low (i.e. "desert resource" according to Combs et al. 2018a). This is plausible in Senegal where the house mouse is only known in villages and towns (Dalecky *et al.* 2015). This situation is well illustrated by the sampling site OKM, which is located in an urban district surrounded by highly vegetated areas (Supplementary information Fig. S2). Although this sampling site

has an estimated N_E equivalent to those found for sampling sites located in area of high genetic continuity (e.g. PFD and POR, Table 1), suggesting favorable habitats for the mouse, OKM also corresponds to the main area of significant genetic discontinuity detected by MAPI (Fig. 1B). The sampling site PAH, located in the zoological park of Hahn, can also illustrate how the low quality of vegetated habitats can impact the population dynamics of the house mouse. As can be expected in unfavorable habitat, this site has one of the smallest estimates of N_E (Fig. 1B, Supplementary information Fig. S2, Table 1). However, PAH is also embedded in the area of high genetic connectivity corresponding to the large industrial/commercial zone running along the east cost of the peninsula from the harbor (Supplementary information Fig S2; Borderon et al. 2014). The highly vegetated zoological park could thus be a local sink that is regularly recolonized by mice emigrating from surrounding areas where population size are much larger (such as from the neighboring sampling site HPE, which is not overly differentiated from PAH: Table 1, Supplementary Information, Table S3).

Among the other urban habitats considered in the CAR model, high standing residential areas (class "Urban 4") had a positive effect, while regular spontaneous habitat (class "Urban 1") had a negative effect on the average level of genetic differentiation. Urban habitat of type 1 is very composite, including medium-high to very poor socio-economics status (Supplementary information Table S1) and characterised by the regular presence of traditional houses built with banco (i.e., mud brick), wood and thatched roofs that are known to be more favorable to rodents than concrete-built houses as they can easily burrow and reach food storages (Makundi *et al.* 1999, Bonwitt *et al.* 2017). Poor building and sanitary conditions have already been identified as critical factors favoring local rodent infestation (Taylor *et al.* 2008, Cavia *et al.* 2009, Himsworth *et al.* 2014).

We acknowledge that the low sampling size and spatial coverage (i.e. clumped sampling of 14 localities, 13 in the Cap-Vert peninsula) prevent us from using individual-based analyses that would be more suited to investigate small scale variations in genetic patterns and identify cityscape features associated to local increase or decrease in genetic connectivity (Richardson et al. 2017, Combs et al. 2018a). However, despite these limitations, the use of the MAPI method, which is spatially explicit and

free from population-model assumptions (Piry et al. 2016), allows to highlight two important points regarding population genetic studies that aim at providing information to help designing control strategies against urban commensal rodents. First, as already pointed out in previous studies, the historical dynamics of urbanization can still be reflected in current genetic patterns (Gardner-Santana *et al.* 2009, Harris *et al.* 2016, Lourenço *et al.* 2017). Not considering the potential effects of past demographic events when studying urban populations of commensal rodents, especially invasive species such as the house mouse, can bias conclusions on current genetic connectivity and then be misleading for the design of control strategies. Second, socio-economic characteristics should be considered along with other land cover variables to characterize cityscapes in urban landscape genetic studies of commensal rodents. Indeed, this information constitute a relevant proxy of spatial variations in habitat quality in terms of resources availability that can impact local rodent population size and then genetic structure (Masi *et al.* 2010, van Adrichem *et al.* 2013, Lucaccioni *et al.* 2016, Lambert *et al.* 2017, Santos *et al.* 2017, Combs *et al.* 2018b, Byers *et al.* 2019). Identification of genetic discontinuities in landscape genetic studies is very often interpreted as an evidence for the presence of a barrier to individual dispersal while spatial structure can actually result from variation in effective population size, i.e. demographic barrier (Berthier *et al.* 2005, Piry *et al.* 2016, Richardson *et al.* 2016). Such a demographic process can be especially relevant for commensal rodents as natural dispersal distances (i.e. not human-mediated) of these species is generally short ranged (< 1 km) and mainly prompted by the lack of feeding and harborage sites (Pocock *et al.* 2005, Byers *et al.* 2019). In a large and spatially continuous population, gene flow will occur through step-by-step migration over generations (IBD). Conversely, for an equivalent migration rate in a spatially discontinuous population, important gaps in densities can generate genetic discontinuities (Berthier *et al.* 2005, Berthier *et al.* 2006).

In term of rodent management, our results for the house mouse in Dakar suggest that areas such as the harbor and the large contiguous industrial/commercial area running north likely sustain high densities and should be considered as a highly connected system when implementing control strategies. These areas may well be a main source of infestation of the southern city blocks where estimates of local effective population sizes and genetic connectivity are quite high (see PRB and PFD in Table 1 and Fig

1B) despite better conditions in terms of socio-economics status and urban infrastructures. In the same line of thinking, city blocks which are surrounded by large vegetated areas (e.g. area around NGR and OKM) correlating with genetic discontinuities could constitute spatial units for mouse control actions. Highly vegetated areas are likely to be “desert resources” habitats and then local density sinks for the house mouse, which suggests that following eradication actions the probability of recolonization though natural dispersal could be relatively low.

CONCLUSION

In this study, we showed that historical and current heterogeneity of urban landscape can reflect in spatial genetic structure of house mouse populations. As suggested by our study, for strict commensal species, population demographic history is likely to be strongly related to the historical development of cities. Including temporal information on urban planning in landscape genetic studies of such species could help in determining to what extend the observed genetic structure is relevant to assess contemporary processes of interest for management. Sampling rats and mice in urban environments is a very challenging task for many reasons: negotiating access to premises from private owners or city agencies, consequences of rodent infestation evidence for business, protecting material from theft, etc. (Parson et al. 2017). For these reason sampling size and coverage are often limited as in the present study, which make it difficult to precisely analyse the relationships between genetic patterns and urban features, especially the impact of the socio-economic environment. In this context, collaborating with both public and private pest control services could help in achieving appropriate sampling design for genetic studies aiming at informing control actions (Parsons et al. 2017).

Acknowledgements

We thank all the people in Senegal who welcomed us into their homes for the purposes of small mammal trapping, as well as the Senegalese Head Office of Waters and Forests, the Ministry of Health

and Social Action in Senegal, the National hygiene service, and the management director of the Dakar harbour for allowing us to do fieldwork and to export samples. Molecular data were generated at the molecular biology platform of the CBGP and the technical facilities of the LabEx CEMEB. We thank Nathalie Sarr for data capture, and Caroline Tatard for their help in molecular biology, J. Papaïx and T. Opitz for helpful comments on the Bayesian CAR model and J.-M. Duplantier for fruitful discussions on the urbanization history and rodent colonization of Dakar. Funding was provided by the Institut de Recherche pour le Developpement (IRD).

BIBLIOGRAPHY

Angel A, Wanless RM & Cooper J (2009) Review of impacts of the introduced house mouse on islands in the Southern Ocean: are mice equivalent to rats? *Biological Invasions* **11**, 1743–1754.

Aplin KP *et al.* (2011) Multiple geographic origins of commensalism and complex dispersal history of black rats. *PloS One* **6**, e26357.

Ba D, Diaw AT, Leclerc G & Mering C (2013) Analyse de la dynamique spatiale du port de Dakar de 1900 à 2009 : enjeux multi-scalaires et aménagements. *Cybergeo: European Journal of Geography* **635**, 1-30.

Benjamini Y & Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289–300.

Berny P, Esther A, Jacob J & Prescott C (2018) Development of Resistance to Anticoagulant Rodenticides in Rodents. In: *Anticoagulant Rodenticides and Wildlife. Emerging Topics in Ecotoxicology (Principles, Approaches and Perspectives), vol 5.* (eds. N van den Brink *et al.*), pp. 386. Cham, Switzerland.

Berthier K, Galan M, Foltete JC, Charbonnel N & Cosson JF (2005) Genetic structure of the cyclic fossorial water vole (*Arvicola terrestris*): landscape and demographic influences. *Molecular Ecology* **14**, 2861-2872.

Berthier K, Charbonnel N, Galan M, Chaval Y & Cosson JF (2006) Migration and recovery of the genetic diversity during the increasing density phase in cyclic vole populations. *Molecular Ecology* **15**, 2665–2676.

Berthier K *et al.* (2016) Black rat invasion of inland Sahel: insights from interviews and population genetics in South Western Niger. *Biological Journal of the Linnean Society* **119**, 748-765.

Besag J, York J & Mollie A (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1-20.

Blangiardo M & Cameletti M (2015) *Spatial and Spatio‐temporal Bayesian Models with R‐INLA*. John Wiley & Sons, Chichester, UK. 308 p.

Bloom D, Canning D & Fink G (2008) Urbanization and the wealth of nations. *Science* **319**, 772-775.

Bonhomme F & Searle JB (2012) House mouse phylogeography. In: *Evolution of the House Mouse* (eds. M Macholan *et al.*), pp. 278-296. Cambridge University Press, Cambridge.

Bonhomme F *et al.* (2011) Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization. *Proceeding of the Royal Society of London serie B* **278**, 1034-1043.

Bonnardel R (1992) *Saint Louis du Sénégal: mort ou naissance?* 421. L'Harmattan, Paris. 423 p.

Bonwitt J *et al.* (2017) At Home with *Mastomys* and *Rattus*: Human–Rodent Interactions and Potential for Primary Transmission of Lassa Virus in Domestic Spaces. *The American Journal of Tropical Medicine and Hygiene* **96**, 935–943.

Borderon M *et al.* (2014) Qualifier les espaces urbains à Dakar, Sénégal. *Cybergeo: European Journal of Geography* **http://geoprodig.cnrs.fr/items/show/46559**, 1-19.

Brouat C *et al.* (2014) Invasion genetics of a human commensal rodent: the black rat *Rattus rattus* in Madagascar. *Molecular Ecology* **23**, 4153-4167.

Byers KA, Lee MJ, Patrick DM & Himsworth CG (2019) Rats About Town: A Systematic Review of Rat Movement in Urban Ecosystems. *Frontiers in Ecology and Evolution* **7**, 13.

Cavia R, Cueto GR & Suarez OV (2009) Changes in rodent communities according to the landscape structure in an urban ecosystem. *Landscape and Urban Planning* **90**, 11-19.

Chapuis MP & Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24**, 621-631.

Chapuis MP *et al.* (2008) Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles. *Molecular Ecology* **17**, 3640-3653.

Chen M, Zhang HB, Liu W & Zhang W (2014) The Global Pattern of Urbanization and Economic Growth: Evidence from the Last Three Decades. *PLoS ONE* **9**, e103799.

Clegg SM, Degnan SMK, J., Moritz C, Estoup A & Owens APF (2002) Genetic consequences of sequential founder events by an island-colonizing bird. *Proceeding of the National Academy of Sciences, USA* **99**, 8127-8132.

Cohen B (2006) Urbanization in developing countries: current trends, future projections, and key challenges forsustainability. *Technology in Society* **28**, 63–80.

Combs M, Puckett EE, Richardson J, Mims D & Munshi-South J (2018a) Spatial population genomics of the brown rat (*Rattus norvegicus*) in New York City. *Molecular Ecology* **27**, 83-98.

Combs M *et al.* (2018b) Urban rat races: spatial population genomics of brown rats (*Rattus norvegicus*) compared across multiple cities. *Proceeding of the Royal Society of London B* **285**, 20180245.

Cucchi T, Auffray JC & Vigne J-D (2012) On the origin of the house mouse synanthropy and dispersal in the Near East and Europe: zooarchaeological review and perspectives In: *Evolution of the House Mouse* (eds. M Macholán *et al.*), pp. 65-93. Cambridge University Press, Cambridge.

Dalecky A *et al.* (2015) Range expansion of the invasive house mouse *Mus musculus domesticus* in Senegal, West Africa: a synthesis of trapping data over three decades, 1983–2014. *Mammal Review* **45**, 176-190.

Dellicour S *et al.* (2019) Landscape genetic analyses of *Cervus elaphus* and *Sus scrofa*: comparative study and analytical developments. *Heredity* online.

Diallo A *et al.* (2012) Asymptomatic Carriage of Plasmodium in Urban Dakar: The Risk of Malaria Should Not Be Underestimated. *PLoS ONE* **7**, e31100.

Do C *et al.* (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources* **14**, 209-214.

Dobson FS (2007) Gene dynamics and social behavior. In: *Rodent Societies* (eds. JO Wolff and PW Sherman), pp. 163-172. University of Chicago Press, Chicago, USA.

Evanno G, Regnaut S & Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Falush D, Stephens M & Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Feng AYT & Himsworth CG (2014) The secret life of the city rat: a review of the ecology of urban Norway and black rats (*Rattus norvegicus* and *Rattus rattus*). *Urban Ecosystem* **17**, 149-162.

Gardner-Santana LC *et al.* (2009) Commensal ecology, urban landscapes, and their influence on the genetic characteristics of city-dwelling Norway rats (*Rattus norvegicus*). *Molecular Ecology* **18**, 2766-2778.

Goudet J (2001) *FSTAT: a program to estimate and test gene diversities and fixation indices (Version 2.9.3).* Available from http://www2.unil.ch/popgen/softwares/fstat.htm.

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**, 95 - 98.

Haniza MZH *et al.* (2015) Large-scale structure of brown rat (*Rattus norvegicus*) populations in England: effects on rodenticide resistance. *Peerj* **3**, e1458.

Hardy OJ & Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Resources* **2**, 618–620.

Harris DA (2011) *Constructing Dakar: cultural theory and colonial power relations in French African urban development*. PhD thesis, University of British Columbia, Burnaby, USA. 135 p.

Harris SE *et al.* (2016) Urbanization shapes the demographic history of a native rodent (the white-footed mouse, *Peromyscus leucopus*) in New York City. *Biology Letters* **12**, 20150983.

Held L, Schrodle B & Rue H (2010) Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. In: *Statistical modelling and regression structures* (eds. T Kneib and G Tutz), pp. 91-110. Springer-Verlag, Berlin.

Himsworth CG *et al.* (2014) A Mixed Methods Approach to Exploring the Relationship between Norway Rat (*Rattus norvegicus*) Abundance and Features of the Urban Environment in an Inner-City Neighborhood of Vancouver, Canada. *PLoS ONE* **9**, e97776.

Jakobsson M & Rosenberg NA (2007) CLUMPP: a cluster matching and permuting program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.

Johnson MTJ & Munshi-South J (2017) Evolution of life in urban environments. *Science* **358**, eaam8327.

Jones EP, Eager HM, Gabriel SI, Johannesdottir F & Searle JB (2013) Genetic tracking of mice and other bioproxies to infer human history. *Trends in Genetics* **29**, 298 - 308

Kajdacsi B *et al.* (2013) Urban population genetics of slum-dwelling rats (*Rattus norvegicus*) in Salvador, Brazil. *Molecular Ecology* **22**, 5056-5070.

Lambert M, Vial F, Pietravalle S & Cowan D (2017) Results of a 15-year systematic survey of commensal rodents in English dwellings. *Scientific Reports* **7**, 15882.

Lippens C *et al.* (2017) Genetic structure and invasion history of the house mouse (*Mus musculus domesticus*) in Senegal, West Africa: a legacy of colonial and contemporary times. *Heredity* **119**, 64-75.

Loiselle BA, Sork VL, Nason J & Graham CB (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). . *American Journal of Botany* **82**, 1420 –1425.

Lourenço A, Álvarez D, Wang IJ & Velo-Antón G (2017) Trapped within the city: integrating demography, time since isolation and population-specific traits to assess the genetic effects of urbanization. *Molecular Ecology* **26**, 1498-1514.

Lucaccioni H *et al.* (2016) From Human Geography to Biological Invasions: The Black Rat Distribution in the Changing Southeastern of Senegal. *PLoS ONE* **11**, e0163547.

M'bokolo E (1982) Peste et société urbaine à Dakar: l'épidémie de 1914. *Cahier d'études Africaines* **85-86**, 13-46.

Makundi RH & Massawe AW (2011) Ecologically based rodent management in Africa: potential and challenges. *Wildlife Research* **38**, 588-595.

Makundi RH, Oguge NO & Mwanjabe PS (1999) Rodent pest management in East Africa – an ecological approach. In: *Ecologically-based Rodent Management* (eds. GR Singleton *et al.*), pp. 460–476. Australian Centre For International Agricultural Research, Camberra.

Martins TG, Simpson D, Lindgren F & Rue H (2013) Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis* **67**, 68-83.

Masi E *et al.* (2010) Socioeconomic and environmental risk factors for urban rodent infestation in Sao Paulo, Brazil. *Journal of Pest Science* **83**, 228-238.

McKinney ML (2006) Urbanization as a major cause of biotic homogenization. *Biological Conservation* **127**, 2 4 7 –2 6 0.

Meerburg BG, Singleton GR & Kijlstra A (2009) Rodent-borne diseases and their risks for public health. *Critical Reviews in Microbiology* **35**, 221-270.

Munshi-South J, Zolnik CP & Harris SE (2016) Population genomics of the Anthropocene: urbanization is negatively associated with genome-wide variation in white-footed mouse populations. *Evolutionary Applications* **9**, 546-564.

Parsons MH, Banks PB, Deutsch MA, Corrigan RF & Munshi-South J (2017) Trends in urban rat ecology: a framework to define the prevailing knowledge gaps and incentives for academia, pest management professionals (PMPs) and public health agencies to participate. *Journal of Urban Ecology* **3**, 1-8.

Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* **10**, 439-446.

Pélissier P (1983) *Atlas du Sénégal*. Les éditions Jeune Afrique, Paris. 72 p.

Piry S *et al.* (2016) Mapping Averaged Pairwise Information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution* **7**, 1463-1475.

Piry S *et al.* (2018) Fine-scale interactions between habitat quality and genetic variation suggest an impact of grazing on the critically endangered Crau Plain grasshopper (Pamphagidae: *Prionotropis rhodanica*). *Journal of Orthoptera Research* **27**, 61–73.

Pocock MJO, Hauffe HC & Searle JB (2005) Dispersal in house mice. *Biological Journal of the Linnean Society* **84**, 565-583.

Pritchard JK, Stephens M & Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Puckett EE *et al.* (2016) Global population divergence and admixture of the brown rat *Rattus norvegicus*. *Proceedings of the Royal Society B: Biological Sciences* **283**, 20161762.

QGIS Development Team (2018) *QGIS Geographic Information System*. Available from http://qgis.osgeo.org.

Rajabi-Maham H, Orth A & Bonhomme F (2008) Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe. *Molecular Ecology* **17**, 627–641.

Ramachandran S *et al.* (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceeding of the National Academy of Sciences, USA* **102**, 15942–15947.

Richardson JL, Brady SP, Wang IJ & Spear SF (2016) Navigating the pitfalls and promise of landscape genetics. *Molecular Ecology* **25**, 849–863.

Richardson JL *et al.* (2017) Using fine scale spatial genetics of Norway rats to improve control efforts and reduce leptospirosis risk in urban slum environments. *Evolutionary Applications* **10**, 323-337.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure *Molecular Ecology Notes* **4**, 137-138.

Rouhana S *et al.* (2015) *Revue de l'urbanisation : villes émergentes pour un Sénégal émergent*. Washington, D.C. 126 p.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-Statistics under isolation by distance. *Genetics* **145**, 1219-1228.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Rozas J *et al.* (2017) DnaSP v6: DNA Sequence Polymorphism Analysis of Large Datasets. *Molecular Biology and Evolution* **34**, 3299-3302.

Salem G (1981) *Grand Dakar un quartier charnière du Cap Vert*. ORSTOM, Dakar. 120 p.

Santos NJ, Sousa E, Reis MG, Ko AI & Costa F (2017) Rat infestation associated with environmental deficiencies in an urban slum community with high risk of leptospirosis transmission. *Cadernos de Saúde Pública* **33**, e00132115.

Sikes RS, Gannon WI & The Animal Care and Use Committee of the American Society of Mammalogists (2011) Guidelines of the American Society of Mammalogists for the use of wild mammals in research. *Journal of Mammalogy* **92**, 235–253.

Singleton GR, Leirs H, Hinds LA & Zhang Z (1999) *Ecologically based management of rodent pests*. ACIAR Monograph N° 59. Australian Centre for International Agricultural Research, Canberra. 494 p.

Singleton GR, Hinds LA, Krebs CJ & Spratt DM (2003) *Rats, mice and people: rodent biology and management*. ACIAR Monograph series 96. Australian Centre for International Agricultural Research, Canberra. 1-114 p.

Singleton GR, Jacob J, Krebs CJ & Monadjem A (2015) A meeting of mice and men: rodent impacts on food security, human diseases and wildlife conservation; ecosystem benefits; fascinating biological models. *Wildlife Research* **42**, 83-85.

Sinou A (1993) *Comptoirs et villes coloniales du Senegal: Saint-Louis, Gorée, Dakar*. Editions Khartala-Orstom, Paris. 364 p.

Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 479-498.

Suzuki H *et al.* (2013) Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity* **111**, 375-390.

Taylor PJ *et al.* (2008) Understanding and managing sanitary risks due to rodent zoonoses in an African city: beyond the Boston Model. *Integrative Zoology* **3**, 38-50.

Temminck CJ (1853) *Esquisses zoologiques sur la côte de Guinée. 1° Mammifères.* E.J. Brill, Leiden. 256 p.

Tixier G (1965) La succession à la régie des chemins de fer de l'AOF. Problèmes posés par l'apparition de nouveaux Etats. *Annuaire Français de Droit International* **11**, 916-932.

Un-Habitat (2008) *Senegal : Profil Urbain de Dakar*. Urban Profiles Programme des Nations Unies pour les Etablissements Humains, Nairobi, Kenya. 36 p.

van Adrichem MHC, Buijs JA, Goedhart PW & Verboom J (2013) Factors influencing the density of the brown rat (*Rattus norvegicus*) in and around houses in Amsterdam. *Lutra* **56**, 77_91.

Vernière M (1973) Pikine, « ville nouvelle » de Dakar. [Un cas de pseudo-urbanisation]. *Espace géographique* **2**, 107-126.

Weir BS & Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Table 1. Genetic estimates of key statistics within sampling sites of *Mus musculus domesticus* in Dakar, Senegal.

For each sampling site, latitude and longitude are indicated in decimal degrees. The number of house mice genotyped for microsatellite markers (N) and sequenced for the mitochondrial D-loop (n_{mt}) are also indicated. For microsatellite data, the table includes an estimation of the population effective size (N_E), as well as the median pairwise kinship coefficients ρ (Loiselle *et al.*, 1995), mean allelic richness (a_r), Nei's unbiased genetic diversity (H_s), and the genetic estimates of F_{IS}. For mitochondrial data, the table includes D-loop haplotypes names, according to Lippens et *al.* (2017) ($H_{D\text{-}loop}$).

Site	code	latitude	longitude	N	N_E	ρ	a_r	H_s	F_{IS}	n_{mt}	$H_{D\text{-}loop}$
Hann pêcheur	HPE	14.724035	-17.429230	47	65.8 [44.1-114.5]	0.043	4.60 ± 1.21	0.64 ± 0.15	0.16	3	H1
Ouakam	OKM	14.723088	-17.490068	46	36.2 [27.4-49.9]	0.108	4.65 ± 1.52	0.58 ± 0.18	0.20	4	H1
Ngor	NGR	14.748939	-17.517352	31	35.3 [23.1-62.0]	0.129	4.42 ± 1.13	0.57 ± 0.16	0.23	4	H1
Park of Hann	PAH	14.726967	-17.434613	19	10.5 [7.5-14.8]	-0.021	4.67 ± 1.68	0.64 ± 0.17	0.14	–	–
Yoff	YOF	14.757200	-17.473321	31	59.7 [38.0-118.4]	0.054	4.86 ± 1.86	0.60 ± 0.22	0.23	4	H1
Plateau-Rebeuss	PRB	14.672396	-17.442259	54	48.9 [37.2-67.6]	0.030	5.71 ± 1.55	0.64 ± 0.13	0.26	–	–
Plateau- Faidherbe	PFD	14.674027	-17.436326	39	41.0 [31.3-56.3]	0.032	5.52 ± 1.73	0.66 ± 0.14	0.11	4	H2/H16
Port of Dakar	POR	14.683414	-17.430436	29	33.5 [24.9-48.5]	0.001	6.11 ± 2.21	0.67 ± 0.16	0.16	8	H1/H16
Grand Dakar	GDK	14.704387	-17.455356	34	25.9 [19.5-36.0]	0.056	5.09 ± 1.44	0.59 ± 0.18	0.24	–	–

Patte d'Oie	PDO	14.748332	-17.437953	39	19.3 [15.9-23.6]	0.036	5.08 ± 1.77	0.68 ± 0.11	0.30	–	–
Colobane	CBN	14.695814	-17.444924	45	45.4 [34.7-62.5]	0.040	4.84 ± 1.46	0.65 ± 0.12	0.16	–	–
Gueule Tapée	GTP	14.680193	-17.458608	43	24.5 [19.7-30.9]	0.046	4.86 ± 1.36	0.61 ± 0.15	0.30	4	H1/H2/H17
Gorée Island	IDG	14.667140	-17.398320	72	21.7 [15.2-31.4]	0.201	3.16 ± 0.91	0.44 ± 0.19	0.26	9	H1/H18
Sacré Cœur 3 (2014)	DAK	14.717635	-17.463809	28	5.8 [4.3-7.3]	0.077	4.77 ± 1.61	0.58 ± 0.19	0.31	5	H1/H2

Figure Legends

Figure 1. Geographical location of sampling sites (code names in Table 1) and genetic structure of *Mus musculus domesticus* in Dakar resulting from: (**A**) STRUCTURE: results are presented for *K* = 2 with colors in pie charts indicating the proportion of house mice assigned to each genetic groups within each site; the barplot indicates individual ancestry estimates (each vertical line represents an individual) in each genetic group; (**B**) MAPI with a color scale representing the spatial variation of the average level of genetic differentiation (F_{ST}) between sampling sites, from lowest values in red to highest values in blue; significant areas of genetic continuity and discontinuity are delineated by dotted and full black lines, respectively, and sampling sites are represented by a circle with a size depending on the spatial distribution of individuals sampled within each site . Small inserts representing the location of Senegal and of Dakar were generated using data from Natural Earth (public domain).

Figure 2. Dakar cityscape. Each map shows the grid used in the MAPI analysis with each cell exhibiting the proportion of one of the seven land cover classes ("Urban 1" to "Urban 6" and "Vegetation" retrieved from Borderon *et al*. (2014)) or the value of the "built-up" (first year of building) and "connection" (year of connection to the first European settlement) variables.

Figure 3. Posterior marginal distributions of the variables considered in the Bayesian CAR model. The posterior probability (pp) to have a significant positive effect is indicated for each variable (i.e. considering a threshold of 0.05, pp > 0.95 indicates a significant positive effect while pp < 0.05 indicates a significant negative effect of the variable).

Figure 1

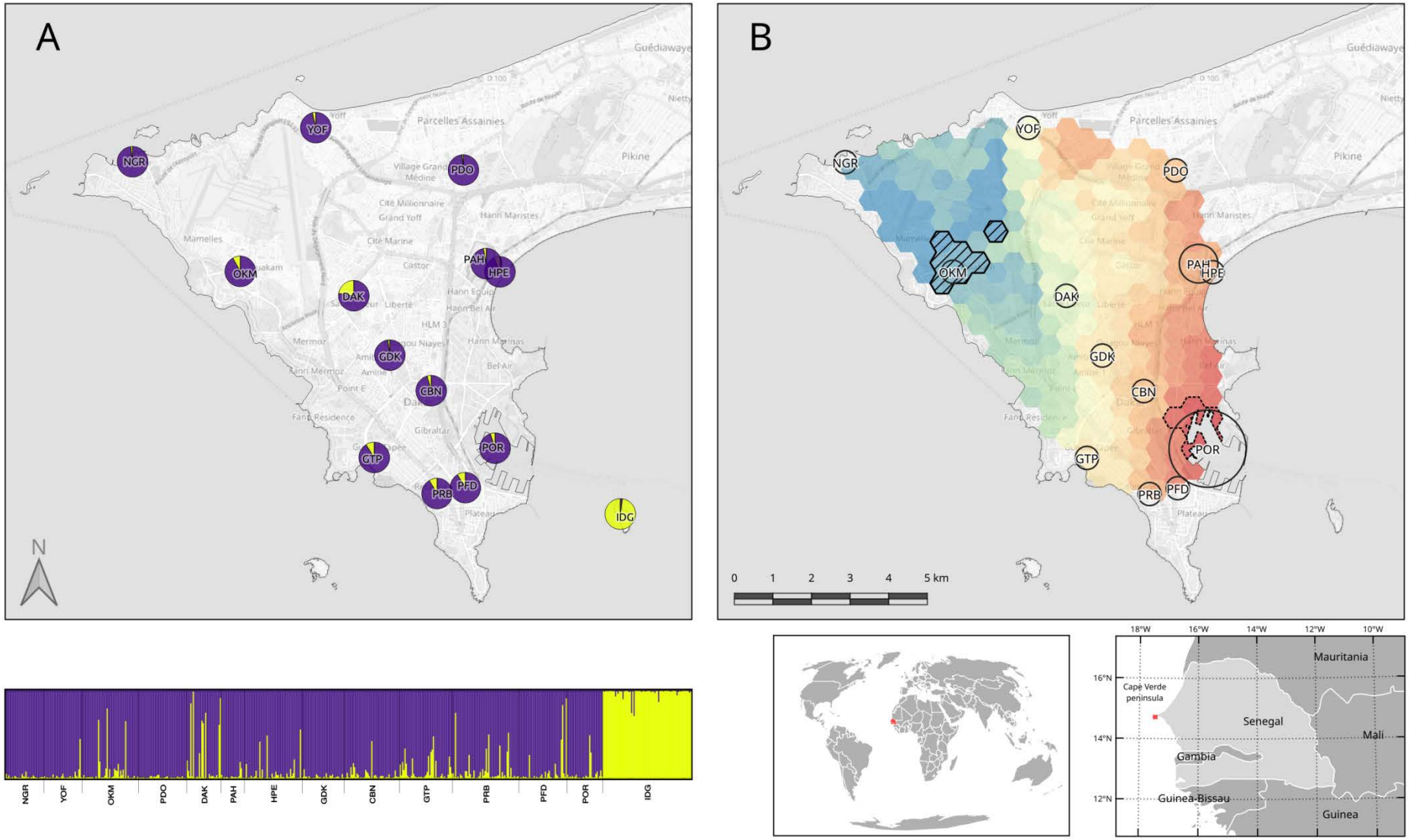

A
B
NGR
YOF
PDO
OKM
DAK
PAH
HPE
GDK
CBN
GTP
PRB
PFD
POR
IDG
N
0 1 2 3 4 5 km
NGR YOF OKM PDO DAK PAH HPE GDK CBN GTP PRB PFD POR IDG
18°W 16°W 14°W 12°W 10°W
16°N
14°N
12°N
Mauritania
Cape Verde peninsula
Senegal
Mali
Gambia
Guinea-Bissau
Guinea

Figure 2

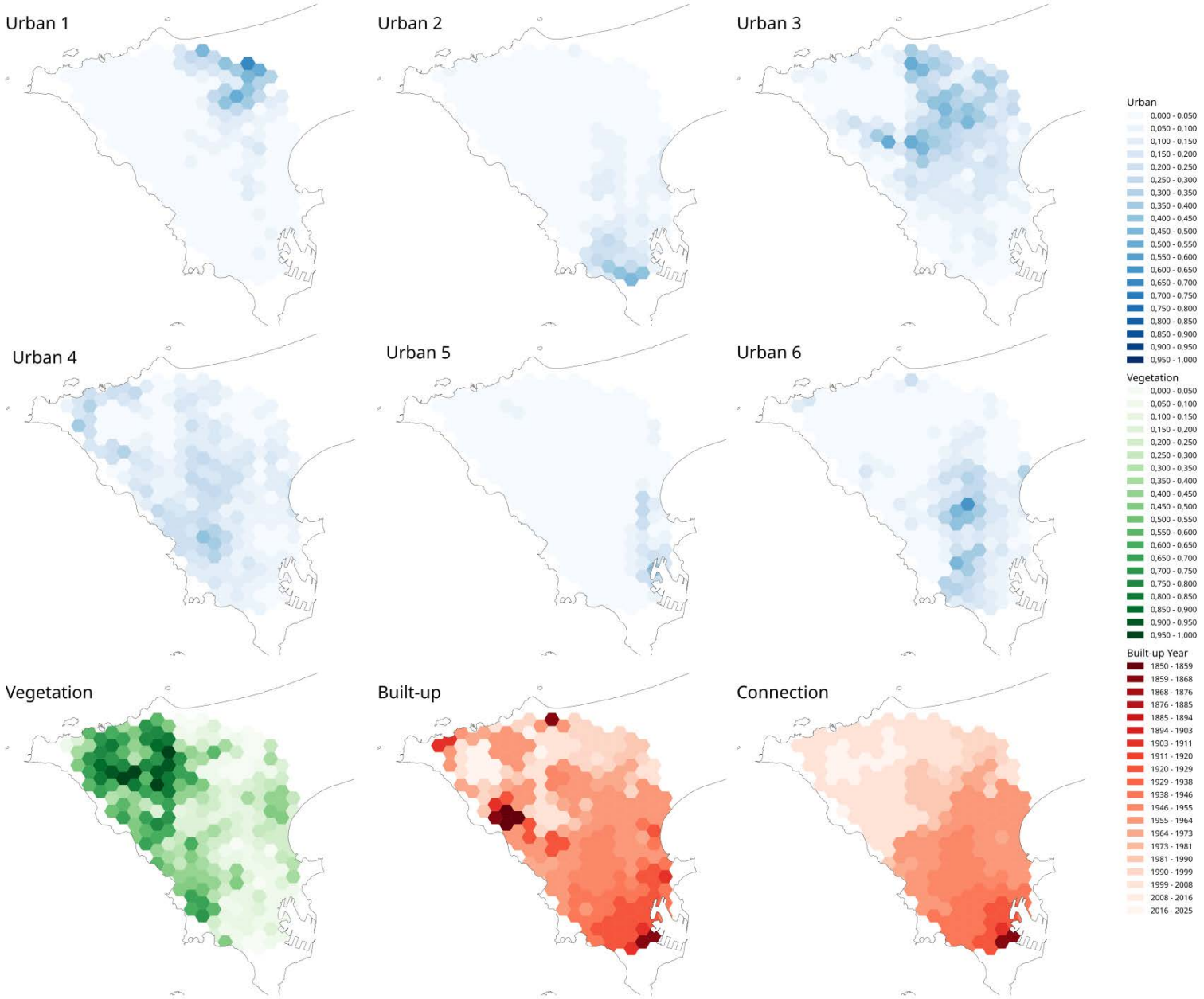
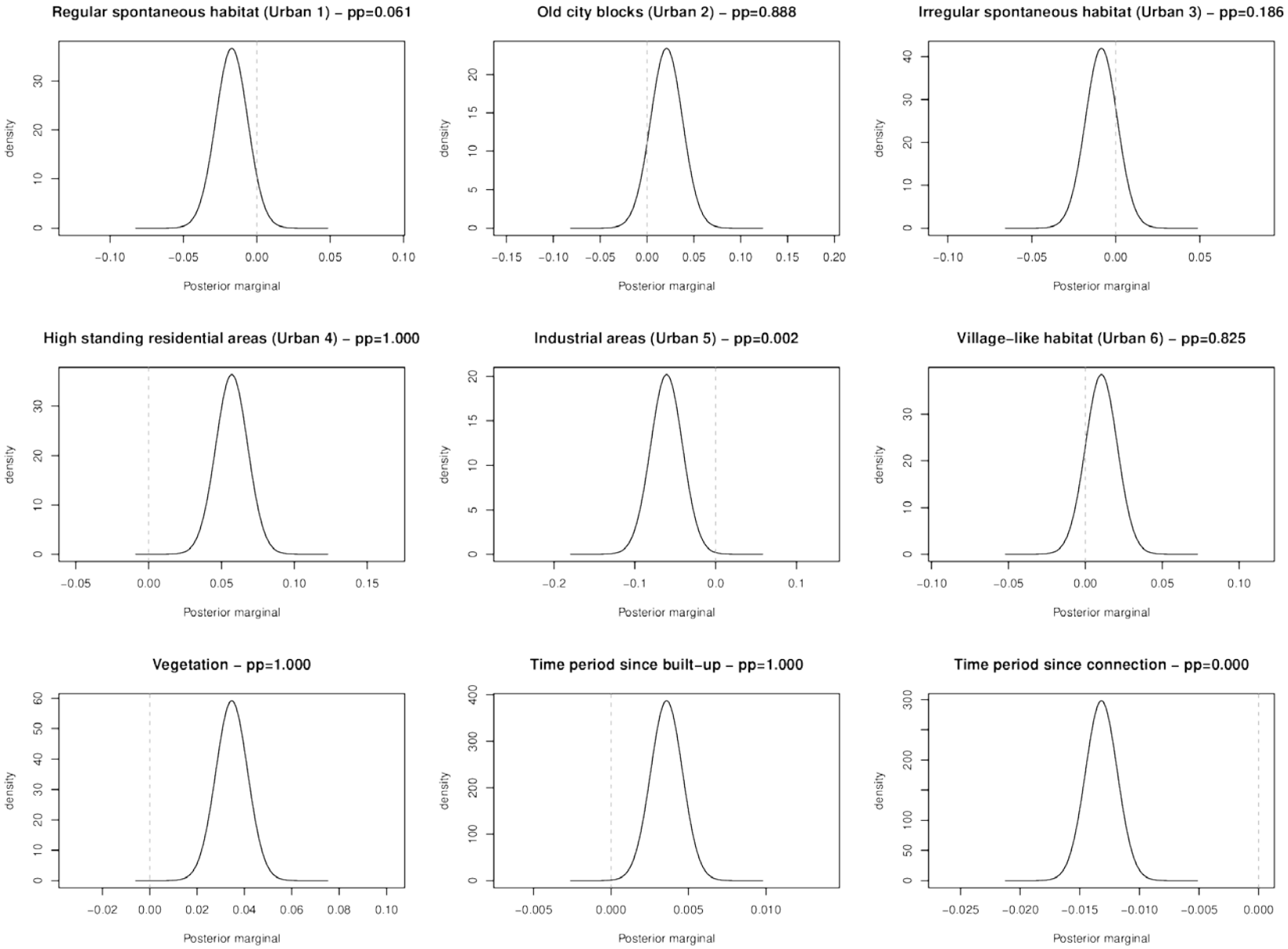

Urban 1
Urban 2
Urban 3
Urban 4
Urban 5
Urban 6
Vegetation
Built-up
Connection
Urban
0,000 - 0,050
0,050 - 0,100
0,100 - 0,150
0,150 - 0,200
0,200 - 0,250
0,250 - 0,300
0,300 - 0,350
0,350 - 0,400
0,400 - 0,450
0,450 - 0,500
0,500 - 0,550
0,550 - 0,600
0,600 - 0,650
0,650 - 0,700
0,700 - 0,750
0,750 - 0,800
0,800 - 0,850
0,850 - 0,900
0,900 - 0,950
0,950 - 1,000
Vegetation
0,000 - 0,050
0,050 - 0,100
0,100 - 0,150
0,150 - 0,200
0,200 - 0,250
0,250 - 0,300
0,300 - 0,350
0,350 - 0,400
0,400 - 0,450
0,450 - 0,500
0,500 - 0,550
0,550 - 0,600
0,600 - 0,650
0,650 - 0,700
0,700 - 0,750
0,750 - 0,800
0,800 - 0,850
0,850 - 0,900
0,900 - 0,950
0,950 - 1,000
Built-up Year
1850 - 1859
1859 - 1868
1868 - 1876
1876 - 1885
1885 - 1894
1894 - 1903
1903 - 1911
1911 - 1920
1920 - 1929
1929 - 1938
1938 - 1946
1946 - 1955
1955 - 1964
1964 - 1973
1973 - 1981
1981 - 1990
1990 - 1999
1999 - 2008
2008 - 2016
2016 - 2025

Figure 3


Regular spontaneous habitat (Urban 1) – pp=0.061
Old city blocks (Urban 2) – pp=0.888
Irregular spontaneous habitat (Urban 3) – pp=0.186
High standing residential areas (Urban 4) – pp=1.000
Industrial areas (Urban 5) – pp=0.002
Village–like habitat (Urban 6) – pp=0.825
Vegetation – pp=1.000
Time period since built–up – pp=1.000
Time period since connection – pp=0.000
density
Posterior marginal