Interplay between historical and current features of the cityscape in shaping 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

Population genetic approaches are often used to investigate dispersal patterns of species living in highly urbanized environment in order to improve management strategies for biodiversity conservation or pest control. However, in such environment, population genetic structure may reflect both current features of the cityscape and urbanization history. This can be especially relevant when focusing on non-native invasive commensal rodents that have been introduced in numerous primary colonial European settlements. Accounting for spatial and temporal cityscape heterogeneity to determine how past and recent demographic events may interplay to shape current population genetic structure of synanthropic rodents 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, where the species may have been introduced as soon as Europeans settled in the middle of the nineteenth century. We examined genetic variation at one mitochondrial locus and 15 nuclear microsatellite markers from individuals sampled in 14 sampling sites representing different stages of urbanization history and different socio-economic environments in Dakar. We used various approaches, including model-based genetic clustering and model-free smoothing of pairwise genetic estimates. We further linked observed spatial genetic patterns to historical and current features of Dakar cityscape using random forest and Bayesian conditional autoregressive models. Results are consistent with an introduction of the house mouse at colonial time and current genetic structure reflects the interplay between the historical dynamics of urbanization and the variation of contemporary urban habitats, including socio-economic features, that translate in habitat quality for the house mouse. Our results highlight the potential importance of accounting for past demographic events to understand spatial genetic patterns of non-native invasive commensal rodents in highly urbanized environment.

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). However, the situation can be quite different for species that 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). Rodent control actions are costly, 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 recognized that the development of more efficient management strategies requires a better understanding of urban rodent evolutionary ecology (Singleton *et al.* 1999, Makundi & Massawe 2011, Feng & Himsworth 2014, Parsons *et al.* 2017, Combs *et al.* 2018b, Byers *et al.* 2019) and notably of variations in gene flow (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, 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). The need to integrate socioeconomic features of the urban environment in population dynamic and genetic studies of commensal rodents is advocated by several studies showing that areas with poor housing conditions (including building material quality, sanitation status within and outside the houses) and commercial premises are more prone to rodent infestations than high-standing residential or touristic areas (Masi *et al.* 2010, van Adrichem *et al.* 2013, Jassat *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 colonial 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 how the interplay between the current and historical features of the cityscape of Dakar, Senegal, may impact 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 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 enable to characterize fine-scale genetic patterns, 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 to which extend considering both the current features of the cityscape and historical information on the urbanization process, may help to explain the spatial pattern of genetic variation of the house mouse in Dakar.

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 (Fig. 1). 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. 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, and (ii) results of a recent classification of Dakar socioeconomic habitats based on satellite images and census data from Borderon *et al.* (2014) to describe the current features of the cityscape.

Historical data were gathered from eleven historical maps and one recent map that were retrieved from various web sources or from paper maps provided by the Senegalese national geographical institute (see 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 (R Core Team 2018) package "sf" (Pebesma 2018). SQL queries on this database allowed to merge adjacent polygons tagged as built-up. We used the city map of 1862 (Fig. 1) to determine the localization of the first European settlement (spatial coordinates: 14.67210°N; 17.429235°W). For each of the 12 maps, the polygons tagged as built-up that overlapped the coordinates of the first European settlement were also tagged as connected. The tags "built-up" and "connected" attributed to the polygons were then used to define two historical variables that can be computed within any kind of spatial unit: (e.g. raster or grid cells, spatial buffers, see below). The first variable, "Built-up", informs on the date at which a spatial unit can be considered as urbanized. This date was determined by considering that built-up polygons from a given historical map must cover at least 5% of the surface of the spatial unit at hand for the latter to be considered as urbanized at the corresponding time point. To allow further analyses, the date of 2025 was attributed to the spatial units that could still not be identified as urbanized from the 2018 map (i.e. less than 5% of the surface covered by "built-up" polygons). The second variable, called "Connection", provides temporal information on the spatial expansion of the city from the first European settlement. As for "Built-up", the date of connection attributed to any spatial unit was based on the historical map for which at least 5% of the surface is covered by polygons tagged as connected. For further analyses, the historical variables were considered 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 spatial unit was considered as urbanized (variable "Built-up") and the year at which a cell was connected to the first European Settlement (variable "Connection").

To describe the current features of the cityscape within the peninsula, we used six socio-economic classes 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 of population wealth, that was related to the urban morphology of Dakar based on classification analyses of SPOT 5 images recorded in September 2007 (see Borderon et al. 2014 for details). These classes, presented in Table 1, distinguish: 1) Regular planned housing of good quality (apartments and high-rise houses) mainly located

in the oldest part of the city and benefiting from infrastructures such as water supply networks, sewers, waste removal; 2) the scattered village-like habitat with a high density of small low-rise houses with no clear delineations; 3) the regular and 4) irregular spontaneous habitats, which correspond to different waves of suburban development between the sixties and the nineties, 5) the more recent development of high-standing residential areas and 6) areas mainly devoted to commercial and industrial activities that have been expanding from the harbor since the first European settlement. As highly vegetated areas may constitute a very unfavorable habitat (desert resource in the sense of Combs et al. 2018a) for the house mouse that is strictly synanthropic in Senegal (Dalecky et al. 2015), we also extracted information on vegetation occurrence to describe the cityscape of Dakar (Supplementary material, section 1, Fig. S1.1). The proportion of each land cover variable was retrieved for each pixel of a 100x100m resolution raster. Due to the successive urbanization policies of Dakar, since the colonial period to these days, some features of the current cityscape are spatially highly imbricated, i.e. they strongly co-vary in space (see section 1 in Supplementary material for historical information on the development of the city). To account for such spatial dependencies between land cover variables, we used a multivariate spatial analysis based on Moran's index. This approach allows introducing a spatial constraint on a Principal Component Analysis (PCA) through the definition of a spatial weighting matrix, i.e. a connection network, used to compute Moran's spatial autocorrelation statistic (MULTISPATI-PCA; Dray et al. 2008). MULTISPATI aims at maximizing the product between the variance and the spatial autocorrelation of the scores rather than just the variance as a classical PCA. Here, we used a minimum spanning tree to define the connection matrix between the 100x100m pixels. The significance of the observed structure was tested using a Monte Carlo test (10000 permutations). Based on both the results of the spatial PCA and historical information, we merged some of the land cover variables (see Results section).

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 2; Fig. 1). They were chosen to represent different stages of

the urban history and different habitats (Table 1 and Supplementary material, section 1). 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 main 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. DAK corresponds to the district of Sacré Coeur founded in 1960 at the heart of the Cap-Vert Peninsula. YOF, OKM, NGR and HPE are inside traditional fisherman villages of Lebou that have been successively absorbed by the urban development after 1960. The sampling site PAH is located within the public and zoological park of Hann.

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, except the Sacré Coeur district (DAK), which was sampled in 2014 for a previous study (Lippens et al. 2017), according to a standardized protocol described in 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). Rodents were brought back to our laboratory, euthanized by cervical dislocation, weighted, sexed and measured the same day. House mouse (easily distinguishable morphologically from

other rodent species) was the most sampled rodent in all sampling site. A piece of tissue (hind foot) for each adult mouse 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 876 bp of 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 2). 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 (731 bp) from this study were aligned with 119 sequences from Senegal (including five sequences from DAK), published by Lippens *et al.* (2017), and with 1958 sequences retrieved from GenBank: 1672 whose references were compiled in Lippens *et al.* (2017) and 168 published in other studies (Storz *et al.*, 2007 ; Förster *et al.*, 2009 ; Castiglia *et al.*, 2005, 2013; Gabriel *et al.*, 2011; Jones *et al.*, 2011, 2012; King *et al.*, 2016; MacKay *et al.*, 2016 ; Garcia-Rodriguez *et al.*, 2018). 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). The haplogroups found in this study have already been found from house mice sampled in Senegal and included in the phylogenetic analysis conducted by Lippens et al. (2017). A phenogram was 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 analysis

Genetic diversity and differentiation

Population genetic analyses were done on a dataset comprising 557 house mice (Table 2: 529 mice sampled for this study; 28 mice from DAK genotyped in Lippens *et al.* (2017) using the same markers). 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 calculated the kinship coefficient (p) (Loiselle *et al.* 1995) between all pairs of individuals within sampling sites, with SPAGeDI v1.4 (Hardy & Vekemans 2002), in order to evaluate whether heterozygote deficiencies in the dataset may be due to the oversampling of closely related individuals or to the pool of individuals sampled from different social group (Parreira & Chikhi 2015). Genotype data from all localities were used as the dataset of reference for allelic frequencies. We also examined whether null alleles may alternatively explain heterozygote deficiencies, by estimating null allele frequencies for each sampling site at each locus using the software FREENA (Chapuis & Estoup 2007).

For each sampling site, genetic diversity was estimated using the expected heterozygosity Hs (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.

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.

Under a model of isolation by distance (IBD), genetic distance between populations is expected to increase with geographical distance (Rousset 1997). IBD was analyzed by regressing pairwise estimates of F_{ST} / (1- F_{ST}) against between-sites log-transformed geographical distances. Mantel tests were performed to test the correlation between genetic differentiation and Euclidean geographical distance using GENEPOP (10 000 permutations). We used the model-based clustering approach implemented in STRUCTURE v2.3.4 (Pritchard *et al.* 2000), in order to estimate the potential 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). IBD and STRUCTURE analyses were performed with and without IDG, in order to analyze the spatial pattern of genetic variation apart from the effect of the geographical isolation of the island.

Spatial patterns of genetic variation and cityscape heterogeneity

The relationship between the spatial patterns of genetic diversity and differentiation and the spatio-temporal urbanization process within the Cap-Vert peninsula (i.e. Gorée Island, locality IDG, was excluded from the dataset) was analyzed using both a punctual (i.e. population-based genetic estimates) and network (i.e. pairwise genetic estimate) approach.

For the punctual approach, we considered three genetic estimates for each sampling site: the allelic richness, the expected heterozygosity and a local-F_{ST} value. The latter was computed using the program GESTE (Foll & Gaggiotti 2006) and represents the average level of differentiation of one site from all the others. To relate these genetic estimates to the characteristics of the cityscape in the surrounding of each sampling site we used nested circular buffers, of radius equal to 300m, 600m, 1000m and 1500m, centered on the barycenter of the individual coordinates at each site. The value for each current (land

cover classes) and historical cityscape variable ("Build-up" and "Connection") was computed as detailed in the section "Spatio-temporal pattern of urbanization". In order to account for collinearity between nested variables, their values within any buffer "i" with a radius superior to 300m corresponded to the difference between the values of the variables computed within the buffer "i" and within the buffer nested within "i" (Marston et al. 2014). For each buffer size, we computed pairwise Spearman correlation coefficients between all variables (genetic estimates and cityscape variables). The relationships between nested cityscape features and local genetic diversity and differentiation indices (a_r, *Hs* and local-F_{ST}) were then explored using Random Forest (RF) approaches (Breiman 2001). We used two different RF algorithms: "ranger" and "cforest" implemented in the R packages "ranger" (Wright & Ziegler 2017) and "party" (Hothorn et al. 2006), respectively. Both approaches allow computing importance value (increase in the mean squared error) for each predictor variable using a permutation procedure. The package "ranger" also provides options for statistical testing of importance values while "party" allows computing conditional permutation importance, i.e. measure the importance of a variable given the other variables in the model. When dealing with correlated variables, conditional permutation importance provides an efficient way to limit the bias toward the systematic selection of variables that are highly correlated to really significant predictors while they may actually be independent to the response variable (Strobl et al. 2008, 2009). For both approaches, RF were performed on the complete dataset, a total of 10,000 trees was built and, as recommended when dealing with correlated predictors, twenty independent replicates of the analyses were done using different values for the initial seed (generated using the random number generator of the R package "random" v. 0.2.6) and "mtry" parameter (ranging from 2 to 14), which determine the number of variables to possibly split at in each node (Strobl et al. 2008). When using "ranger", variable importance and p-values were computed using the permutation (N=10,000) approach of Altmann et al. (2010). The shape of the relationship between genetic estimates and each cityscape feature, computed within the different buffers, was assessed using accumulated local effects (ALE) plot, which are more efficient with highly correlated predictors (Apley 2017). The "cforest" analysis was run by setting the "minsplit" (minimal number of observations for splitting) and "minbucket" (minimal number of observation in terminal nodes) parameters to 6 and 3, respectively. The shape of the

relationships between cityscape features and genetic estimates was assessed using the "partial_dependence" function from the R package "edarf" (Jones & Linder 2017).

For the pairwise approach, we used the MAPI method, which has a low sensitivity to IBD and is not based on a predefined population genetic model (Piry et al. 2016 - implemented in the R package "mapi"). MAPI is a smoothing-like procedure applied on pairwise genetic measures which are attributed to a network of ellipses linking the sampling points. The method produces geographical maps representing the spatial variation in the average level of genetic differentiation between sampled localities. MAPI was run using F_{ST} values computed between all pairs of localities to build the spatial genetic network of ellipses. 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 = 1,000 m) and the Hann park 'PAH' (radius = 500 m). A grid of 243 hexagonal cells with a half-width of 310m 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 high genetic continuity and discontinuity. The MAPI grid was then superimposed on the layers describing the historical and current cityscape and the value of each cityscape variable within each cell of the grid was computed as detailed in the section "Spatio-temporal pattern of urbanization". As for the punctual approach, Spearman correlation coefficient was computed for all variable pairs associated to the cells of the MAPI grid: smoothed-F_{ST} and current and historical cityscape variables. To identify which cityscape features can better explain variations in MAPI smoothed-F_{ST} values we used the conditional RF algorithm implemented in the function "cforest" from the R package

"party". We used different seed and mtry values (from 2 and 5) to perform the analysis (10,000 trees) on 100 independent training datasets built by randomly subsampling 50% of the MAPI grid cells. Result from each training dataset was used to predict smoothed-F_{ST} within the remaining MAPI cells (i.e. not included in the training dataset). Model accuracy was assessed by computing the Pearson correlation coefficient between predicted and observed MAPI smoothed-F_{ST} as well as the R^2 and RMSE (root mean square error).

Finally, we used a conditional autoregressive model (CAR) to further analyze the relationships between MAPI smoothed-F_{ST} values and the cityscape variables computed within each cell. To account for spatial autocorrelation, the error term of the model 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) (specification of the model for R-INLA is provided in Supplementary material, Section 3). Bayesian inference resulted in posterior densities for all predictors from which significance was assessed with a threshold of 0.05. The CAR model was run, first, considering only the land cover classes describing the current cityscape. We then added, one at a time, the historical variables "Built-up" and "Connection" and computed the weighted Akaike Information Criteria (wAIC) as well as the Deviance Information Criteria (DIC) to determine whether adding historical information on the urbanization process allows to better explain the genetic structure of the house mouse than when considering only the current features of the cityscape. To evaluate the fit of the model to the data we computed the Pearson correlation coefficient (r) between predicted and observed values as well as the 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.

MAPI is a smoothing procedure that performs better when the sampling maximizes the spatial coverage of the study area (Piry et al. 2016). To evaluate whether the observed genetic structure can be biased, due to the smoothing of a relatively small network (i.e. clumped sampling of 13 localities in the Cap-Vert peninsula), we applied the approach combining the MAPI analysis and Bayesian CAR model on: 1) individual microsatellites genotypes simulated under a landscape configuration corresponding to a spatial gradient from a favorable to an unfavorable habitat as fully described in Piry et al. (2016), and 2) subsampling of the initial dataset, with a number of sampling sites (N) varying from 6 to 12 (see Supplementary Material, section 4, for details on the simulation and subsampling framework).

RESULTS

Spatio-temporal pattern of urbanization

The Monte-Carlo permutation test of the MULTISPATI-PCA performed on the seven initial land cover variables describing the current cityscape at a 100x100m resolution was highly significant ($p<0.001$). The first two axes of the spatial PCA explained about 50% of the variance. The first axis strongly reflected the opposition between the occurrence of highly vegetated areas in the western part of the city and the very dense urbanized areas in the eastern part. The second axis showed the strong spatial correlation between the land cover classes “Old city blocks” and “Village-like” on the positive side and between the classes representing the “Regular spontaneous” and “Irregular spontaneous” habitats on the negative side (Supplementary material, Section 1, Figs. S1.1, S1.2 and S1.3, Tables S1.1 and S1.2). Based on the results of both the spatial PCA and historical information (see Supplementary material, section 1), we merged the spatially imbricated land cover classes ‘Old city blocks” and “Village-like” into a single class called “Old city” and the classes “Regular spontaneous” and “Irregular Spontaneous” into a same class called “Spontaneous”. Five variables (“Old city”, “Spontaneous”, “Residential”, “Industrial” and

"Vegetation" were then retained to characterize the current cityscape in further analyses (see Table 1 and Fig. 1).

Mitochondrial sequence analysis

There were nine variable sites defining five haplotypes (mean haplotype diversity: 0.264 ± 0.313; mean nucleotide diversity: 0.003 ± 0.005; Fig. 2; Table 2) among the 45 D-loop sequences of house mice from Dakar (40 sequences from this study, five sequences from Lippens et al. 2017). 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). Both haplotypes had been reported at relatively high frequencies in different part of the world, notably Western Europe (Supplementary material, section 2, Table S2.1). 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(Fig. 2A). 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), and in some localities from Western Europe and Canaries Islands (Supplementary material, section 2, Table S2.1). Haplotype diversity was greater in sampling sites of the southern part of the Cap-Vert Peninsula (Table 2; Fig. 2A).

Microsatellite analyses

Genetic diversity and differentiation

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 (Table 2; Supplementary material, section 2, Table S2.2). 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) to not affect significantly estimates of genetic diversity and differentiation (Chapuis & Estoup 2007, Chapuis *et al.* 2008). Hence, this locus was conserved for further 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 IDG (Table 1). Distributions of *p* within sites were largely unimodal (Supplementary material, section 2, Fig. S2.1) and globally zero-centred. This suggests that the observed deviations from Hardy-Weinberg equilibrium result from the presence, within each site, of several social groups from which a few individuals have been sampled, rather than a strong family effect due to the oversampling of highly related individuals from 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). 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 material, section 2, Table S2.3). 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]).

The program STRUCTURE was used to determine whether house mice were spatially structured as well-defined genetic units in Dakar. The highest delta$K$ value was obtained for K = 2 (delta$K$ = 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. 2B). 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. Results were similar (no clear pattern of spatial structure within the Cap-Vert peninsula) when IDG was excluded from the analysis.

Spatial patterns of genetic variation and cityscape heterogeneity

Posterior mode of local F_{ST} estimates, computed for each sampling site with GESTE, ranged from 0.023 ($HDPI_{95\%}$ = [0.011 ; 0.038]) to 0.130 ($HDPI_{95\%}$ = [0.097 ; 0.171]) in POR and OKM, respectively (Supplementary material, Section 3, Table S3.1). The spatial pattern of local genetic differentiation exhibited a gradient-like structure, from the lowest local-F_{ST} values estimated for the sampling sites located in the South part (POR, PFD, PRB) of the Cap-Vert peninsula to the highest local-F_{ST} values estimated for the sampling sites located in the North-Western part (OKM, NGR) (Fig. 2C).

Cityscape features identified by the "ranger" and "cforest" RF methods as the most important to explain a_r, H_s and local-F_{ST} values were very consistent whatever the setting of the analysis (i.e. variation of the" mtry" value from 2 to 13 with 20 replicates using different initial seeds). Moreover, for both the unconditional and conditional RF approaches, there was no evidence for a bias toward the selection of highly correlated predictor variables (Supplementary material, section 3, Fig. S3.1 and Table S3.2) during tree building, i.e. systematic ranking of highly correlated variables among the most important (Strobl et al. 2008, 2009). The "ranger" method, which provides a statistical testing option, identified seven nested cityscape variables as significant to explain the variation of a_r (Supplementary material, section 3, Fig. S3.1A). Among these variables, four corresponded to the historical variable "Connection" computed within the different buffers, two to the variable "Vegetation" computed within the 600m- and 1000m radius buffers and the last one to "Built-up" from the 1000m radius buffers. The same seven variables were also identified as significant to explain the variation in local-F_{ST} values (Supplementary material, section 3, Fig. S3.1C). These seven variables were also ranked as the most important to explain both a_r and local-F_{ST} values when using "cforest". As could be expected considering that a_r and local-F_{ST} were strongly negatively correlated (Supplementary material, section 3, Table S3.2), both the ALE and partial dependence plots showed an opposite effect of "Connection" on a_r and local-F_{ST}, regardless of the buffer size (i.e. increase of a_r and decrease of local-F_{ST} - Supplementary material, section 3, Figs. S3.2A and

S3.2B). In the same way, the variable "Vegetation" at 600m and 1000m had a negative and positive effect on a_r and local-F_{ST}, respectively. Finally, "Built-up" at 1000m had an opposite effect, i.e. increase and decrease of a_r and local-F_{ST}, respectively. Although *Hs* and a_r were positively correlated (Supplementary material, section 3, Table S3.2), both RF methods identified different variables to explain the variation of *Hs*. Significant variables from "ranger" included the variables "Industrial" computed within the 600m-, 1000m- and 1500m-radius buffers, the variable "Vegetation" from the 1000m- and 1500m radius buffers and the variable "Residential" from the 1000m-radius buffer (Supplementary material, section 3, Fig. S3.1B). These variables were also identified among the most important variables when using the conditional "cforest" approach. Both ALE and partial dependence plots suggested a positive effect of "Industrial" on *Hs* while a negative effect was identified for the variables "Vegetation" and "Residential" (Supplementary material, section 3, Figs. S3.2A and S3.2B).

Similarly to GESTE, the results of the MAPI analysis clearly showed a gradient-like pattern from lowest smoothed-F_{ST} values in the eastern part of the study area to highest values in the western part (Fig. 2D). A significant area of genetic discontinuity encompassed the northern part of the Ouakam district (area around the sampling site OKM), between the two main runways of the Léopold Sédar Senghor Airport, while a significant area of high genetic continuity (i.e. lowest smoothed-F_{ST} values) was found in the harbor area close to the first European settlement. Spearman correlation tests showed that MAPI smoothed-F_{ST} values were positively correlated to the variable "Vegetation" and "Residential (R_s = 0.65 and 0.16, respectively) and negatively correlated to all other variables (R_s varied from -0.32, for the variable "Spontaneous", to -0.65 for the variable "Connection" – Supplementary material, section 3, Table S3.3). The ranking of the conditional permutation importance of the cityscape features from the "cforest" RF trained on 100 randomly resampled datasets (50% of the MAPI grid cells) was very consistent when using different seed and "mtry" values": 1) "Connection", 2) "Vegetation", 3) "Industrial", 4) "Spontaneous", 5) "Residential", 6) "Old_City" and 7) "Built-up" (Supplementary material, section 3, Fig S3.3 and S3.4). The predictive accuracy of the model was relatively high and slightly varied when increasing the "mtry" value from 2 to 5. Pearson correlation coefficient, R^2 and

RMSE, averaged over the 100 replicates, ranged: from r = 0.81 ± 0.029 to r = 0.83 ± 0.030; from R^2 = 0.59 ± 0.034 to R^2 = 0.68 ± 0.046; from RMSE = 0.014 ± 0.001 to RMSE = 0.012 ± 0.001 for "mtry" = 2 and "mtry" = 5, respectively.

For all CAR models applied on MAPI smoothed-F_{ST}, the Pearson correlation coefficient between predicted and observed values was ≥ 0.95 and 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 in any case. Moreover, in all cases, the distribution of the PIT values did not significantly depart from a uniform distribution (Kolgomorov-Smirnov test, p-value > 0.05; Table 3), which confirmed that the model was well calibrated. Model residuals were not overly structured in space (Supplementary information, section 3, Fig S3.5). Results (Table 3) showed that considering historical information on the urbanization process allows to better explain the genetic structure of the house mouse in Dakar than current cityscape features alone. Based on wAIC and DIC values, the best improvement was observed when the historical variable "Connection" was included as a covariate (model 3). Under model 3, a significant (5% threshold) positive effect was observed for the variable "Vegetation" and "Residential" (i.e. the higher the frequency of this classes, the higher the average level of genetic differentiation) while a significant negative effect was observed for the variable "Connection", "Industrial" and "Spontaneous". Considering both historical variables (model 4) only slightly improved the model. The variables "Built-up" and "Spontaneous" showed a change in posterior probability value across the different models, which depended on the presence of the variable "Connection" among the predictors.

When combining MAPI and a CAR model to analyze the 200 simulated genotypes datasets (Supplementary material, Section 4, Fig S4.1), we found the expected significant positive effect of the unfavorable habitat on the MAPI smoothed-F_{ST} for 95.5% of the simulated datasets. None of the remaining datasets showed an unexpected significant negative effect. Moreover, the analysis of subsampled datasets (with a number of sampling sites N varying from 6 to 12) showed that the gradient-like pattern in smoothed-F_{ST} and the main effects of the cityscape variables were observable even with N

= 6 sampling sites: high posterior probability for a negative effect of “Connection” and “Industrial” and positive effect of “Vegetation”, “Residential” and “Built-up” (Supplementary material, Section 4, Fig S4.2). For the variables “Old_city” and, more especially, “Spontaneous”, the number of sampling sites sampled was quite crucial to obtain consistent posterior probabilities across the replicates.

DISCUSSION

In this study, we aimed to investigate 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 may be reflecting both the heterogeneity of the current cityscape and the historical dynamics of urbanization. We used various approaches, including population genetics model-based clustering, model-free smoothing of genetic differentiation as well as random forest algorithms that allow investigating non-linear relationships and provide options to account for multicollinearity between predictor variables (i.e. conditional permutation importance) and Bayesian conditional autoregressive models, which allow accounting for spatial autocorrelation. We also performed a simulation study and an analysis of subsampled datasets to assess whether the observed genetic structure could be biased due to the small sampling size (i.e. clumped sampling of 13 localities in the Cap-Vert peninsula).

Beyond the well-marked genetic differentiation of Gorée Island from the Cap-Vert Peninsula, there was no evidence that house mice were structured as delineated genetic units in Dakar. Moreover, within the Peninsula, patterns of mitochondrial genetic variation strongly suggested an introduction of the house mouse during colonial time. Indeed, all mitochondrial haplogroups found in Dakar are typical from Western Europe (Supplementary material, section 2, Table S2.1; Bonhomme *et al.* 2011) with the highest level of diversity observed in sampling sites at the south of the Cap-Vert peninsula (Table 2, Fig. 1). These results are consistent with the historical records of rodent fauna that reported large populations of mice in colonial coastal cities of Senegal from the middle of the 19th century (Temminck 1853), which

corresponds to the foundation of the first European settlement in the harbor area at the south of the Cap-Vert Peninsula.

Spatial expansion of introduced populations may be characterized by sequential founder events, leading to spatial decrease of genetic diversity and increase of genetic differentiation along colonization axes (Clegg *et al.* 2002). For the house mouse in the Cap-Vert Peninsula, analysis of both punctual and pairwise genetic estimates showed that, the shorter the period elapsed since the connection of a built-up area to the first European settlement, the lower the level of allelic richness and the higher the level of genetic differentiation. These results suggest that once the house mouse was introduced within the Peninsula, its expansion followed the spatial axes of continuous urban expansion from the first European settlement. This result is consistent with the ecology of the house mouse, which is known to be strictly commensal in Senegal (i.e. confined to human buildings, Dalecky *et al.* 2015) and relies on human settlements to colonize new habitats. In line with such a strict commensalism, the punctual and pairwise approaches showed strong negative and positive relationships between the proportion of vegetation and the level of genetic diversity (*ar*, *Hs*) and genetic differentiation (local-F_{ST} and smoothed-F_{ST}), respectively, suggesting that highly vegetated areas (i.e. no or a few buildings used for housing, commercial activities…) act as desert resource habitats and/or dispersal barriers for this species.

Beyond the variable "Vegetation", RF analyses on punctual genetic estimates identified a different set of important variables to explain *ar* and local-F_{ST} in one hand and *Hs* in the other hand. Indeed, the historical variables, especially "Connection" (all buffer sizes) and more marginally "Built-up (1000m buffer), had a strong effect on *ar* and local-F_{ST} while *Hs* was mostly impacted by the urban land cover "Industrial" and, to some extent, "Residential". Variation in local a_r estimates depends on the presence/absence of the different alleles sampled through the whole metapopulation while changes in H_s are more dependent on allele encounter probability (mating between individuals carrying different alleles) and, then on the evenness of common allele frequencies (Nei 1975). Reduction in local effective population size (e.g. founder effect) has then a strong and direct impact on a_r due to the rapid loss of rare alleles that already seldom contribute to H_s (Greenbaum et al. 2014). The fact that a_r and local-F_{ST} were

better explained than H_s by historical variables (especially "Connection") may result from the historical isolation and loss of original rare alleles due to founder events during the expansion process (Comps et al. 2001, Roberts & Hamann 2015). In contrast, H_s would be better explained by current features of the cityscape that are likely to impact local density of house mouse.

The disparity in the set of variables identified as the most important for H_s and for a_r or local-F_{ST} was relieved when applying the pairwise approach, which consider between sites spatial heterogeneity of both genetic differentiation (i.e. MAPI smoothed-F_{ST}) and cityscape. The positive effect of the variables "Vegetation" and "Residential", as well as the negative effect of the variables "Connection" and "Industrial" were retrieved from all analyses methods performed to explain the spatial variation in MAPI-smoothed-F_{ST} values (i.e. Conditional Random Forests and Bayesian CAR models). These results further suggest that the observed gradient in genetic differentiation still reflects the historical process of continuous expansion of the city within the Peninsula (Gardner-Santana *et al.* 2009, Harris *et al.* 2016, Lourenço *et al.* 2017) and that deviations from this main pattern would depend on current variations in habitat quality in terms of resources availability (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). Here, the higher level of genetic diversity (H_s) and lower level of genetic differentiation in industrial areas suggest that this is the most favorable habitat to sustain large populations of mice, which is very likely considering that these areas include numerous warehouses to stock different food products such as groundnuts, rice, sugar and flour (Ba *et al.* 2013). In contrast, lower genetic diversity (H_s) and higher genetic differentiation in high-standing residential areas suggest that this urban habitat is the less prone to mouse infestations, as already reported for rats (Masi *et al.* 2010, van Adrichem *et al.* 2013, Lambert *et al.* 2017, Combs *et al.* 2018b, Byers *et al.* 2019). Spatial variations in the level of genetic differentiation are often interpreted as an evidence for the presence of physical barriers to individual dispersal while spatial structure can actually mainly 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 very 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). These two situations could correspond to house mouse dynamics in industrial areas for the former and in high-standing residential and, as an extreme case, highly vegetated areas for the latter.

Result from the RF and CAR models also suggested that the land cover class "Spontaneous" had a negative impact on genetic differentiation. In the study area, the "Spontaneous" urban habitat mainly corresponds to unplanned housing areas (i.e. irregular spontaneous habitat, see Table 1 and Supplementary material, Section 1, Fig. S1.2). These areas, which are occupied by Dakar lowest income households, often lack basic urban infrastructures (e.g. waste removal) and houses are mainly built with precarious material such as wood, metal sheet or banco (i.e. mud brick). Such building materials are already known to be more favorable to rodents than concrete-built houses, as they can more easily circulate, burrow and reach food storages (Makundi *et al.* 1999, Bonwitt *et al.* 2017). Then this urban habitat is expected to be favorable for the house mouse. However, as suggested by the results of the CAR models on the initial and subsampled datasets, retrieving a clear effect for this habitat would depend on both the other predictors included in the model (Table 3) and the number of sites sampled (Supplementary material, Section 4, Fig. S4.2). As CAR models account for spatial autocorrelation (i.e. local smoothing of estimates over neighbouring areas), local effect may emerge to explain a part of the variation not well handled by the strongest predictors of the global pattern. Here, local variation in the pattern of genetic differentiation may be due to historical events related to the "Spontaneous" urban habitat between the sampling sites YOF and PDO. Results from the MAPI and GESTE analyses show that genetic differentiation is quite low in this area located along a district called "Parcelles Assainies", which has been created in 1974 to reduce the population in the city center by encouraging the displacement of the less fortunate into this new borough (Ndiaye 2015; Fig. 1). Human-mediated dispersal of house mice

during these movements from the old city to this peripheral suburb could explain the relatively low level of differentiation between YOF and PDO and then the negative effect associated to the “Spontaneous” habitat. Such a local effect would explain that, though this habitat is present in various sectors of the city, detecting any clear trend from the resampled datasets requires a sampling scheme including the sites bordering the “Parcelles Assainies” (i.e. YOF and PDO). Local effects can also explain the change in the posterior probability observed for the variable “Built-up” according to the presence of the variable “Connection” among the predictors (model 2 and model 4, Table 3). “Built-up” did not appear as a strong predictor when it was the single historical variable in the model (model 2). This was supported by the result of the RF analysis, which despite the high level of correlation between “Built-up” and “Connection”, clearly identified the former as the less important predictor to explain the global pattern (Supplementary material, Section 3, Fig. S3.3). However, when a strong historical predictor was included in the model (“Connection”, model 4, Table 3), the variable “Built-up” exhibited a significant positive effect on genetic differentiation, which is likely to be due to local effects: presence of Lebou villages, which were clearly identified as built-up areas in early maps (see map of 1862 in Supplementary material, section 1, Fig S1.1) while they have been among the last areas to be connected to the first European settlement connected (see maps of 1993 and 2001). These villages have now been absorbed in the city, however there are associated to high levels of differentiation (Fig 1, Fig 2 and Table S2.3 in Supplementary material, section 2). This further suggests that even if the house mouse has been introduced in these villages quite early, they have been quite isolated and historical gene flow intensity across the peninsula has been primarily dependent on the continuous spatial expansion of the city.

Sampling rodents in urban environments may be a challenging task, notably because it requires negotiating access to premises from private owners or city agencies (Parson et al. 2017). For such a reason, sampling size and coverage may be limited which make it difficult to precisely analyse the relationships between genetic patterns and urban features. In this study, despite the small sampling size (i.e. clumped sampling of 13 localities within the peninsula), the analysis of simulated and subsampled datasets showed that the MAPI method allowed retrieving the main genetic structure (i.e. gradient like

pattern) of the house mouse within the peninsula and to relate it to important historical and current features of Dakar cityscape. However, we do believe that, in this particular case, the approach was facilitated by the strength and simplicity of the main spatial structure (i.e. gradient like pattern with a high level of genetic differentiation). More complicated or fine-scale genetic structure would be hard to detect without further increasing the sampling size and spatial coverage through a more intensive individual-based sampling design within and across urban habitats (Richardson et al. 2017, Combs et al. 2018a). In term of rodent management, our results suggest that defining spatial units for house mouse control might not be straightforward in Dakar and, as already pointed out for rats in other cities, classical scales such as individual property or city block are likely to be irrelevant in many local contexts (e.g. Gardner-Santana et al. 2009; Kajdacsi et al. 2013). Relevant spatial scale for management will vary across the city due to habitat quality heterogeneity. For example, areas such as the harbor and the large contiguous industrial/commercial area running north are likely to sustain high densities should be considered as a highly connected network when implementing control strategies (Byers et al. 2019). Areas where house mouse groups are likely to be more scattered (Residential areas and highly vegetated environment) could constitute spatial targets for localized control as the probability of recolonization though natural dispersal could be relatively low following treatments.

CONCLUSION

In this study, we showed that the spatial genetic structure of a strictly commensal species such as the house mouse likely result from the interplay between the spatio-temporal process of urbanisation and the current heterogeneity of the cityscape. First, spatial genetic patterns are consistent with an introduction of the house mouse in the Cap-Vert Peninsula during colonial time followed by a spatial expansion driven by the development of the city from the first European Settlement. Second, the genetic structure within the Peninsula mainly corresponds to a gradient-like pattern matching the historical process of spatially continuous expansion of the city. Deviations from this average pattern reflect strong spatial differences in

current urban land cover with highly vegetated areas and residential areas being the less favorable habitat while the industrial areas would be highly favorable to mouse infestation, as already observed for rats.

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. Sampled are preserved in the CBGP-Small mammal collection (https://doi.org/10.15454/WWNUPO) and referenced in the Small Mammal Database associated. Molecular data were generated at the molecular biology platform of the CBGP and through the GenSeq technical facilities, with the support of the LabEx CeMEB. We thank Nathalie Sarr for data capture, Caroline Tatard for helping with molecular laboratory work, J. Papaïx and T. Opitz for helpful comments on the Bayesian CAR model and J.-M. Duplantier for fruitful discussions on urbanization history and rodent colonization of Dakar. Funding was provided by the Institut de Recherche pour le Developpement (IRD).

BIBLIOGRAPHY

ANSD (2016) *Situation économique et sociale du Sénégal en 2013*. Report of the Agence Nationale de la statistique et de la démographie, 48p.

Altmann A, Tolosi L, Sander O & Lengauer T (2010) Permutation importance: a corrected feature importance measure, *Bioinformatics* **26**, 1340-1347.

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.

Apley D (2017) ALEPlot: Accumulated Local Effects (ALE) Plots and Partial Dependence (PD) Plots, 2017. URL https://CRAN.R-project.org/package=ALEPlot. R package version 1.0.

Aplin KP, Suzuki H, Chinen AA, Chesser RT, ten Have J, Donnellan SC, Austin J, Frost A, Gonzalez J-P, Herbreteau V, Catzeflis F, Soubrier J, Fang Y-P, Robins J, Matisoo-Smith E, Bastos ADS, Maryanto I, Sinaga MH, Denys C, Van Den Bussche RA, Conroy C, Rowe K & Cooper A (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, Garba M, Leblois R, Navascués M, Tatard C, Gauthier P, Gagaré S, Piry S, Brouat C, Dalecky A, Loiseau A & Dobigny G (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, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffray J-C & Britton-Davidian J (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.

Bonwitt J, Saez AML, J., Ansumana R, Dawson M, Buanie J, Lamin J, Sondufu D, Borchert M, Sahr F, Fichet-Calvet E & Brown H (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, Oliveau S, Machault V, Vignolles C, Lacaux J- & N'Donky A (2014) Qualifier les espaces urbains à Dakar, Sénégal. *Cybergeo: European Journal of Geography* **http://geoprodig.cnrs.fr/items/show/46559**, 1-19.

Breiman L.(2001) Random forests. Machine Learn **45**(1), 5–32.

Brouat C, Tollenaere C, Estoup A, Loiseau A, Sommer S, Soanandrasana, R., Rahalison L, Rajerison M, Piry S, Goodman SM & Duplantier JM (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.

Castiglia R, Annesi F & Capanna E (2005) Geographical pattern of genetic variation in the Robertsonian system of *Mus musculus domesticus* in central Italy. *Biological Journal of the Linnean Society* **84**, 395-405.

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

Chapuis MP, Lecoq M, Michalakis Y, Loiseau A, Sword GA, Piry S & Estoup A (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, Byers KA, Ghersi BM, Blum MJ, Caccone A, Costa F, Himsworth CG, Richardson JL & Munshi-South J (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.

Comps B, Gomory D, Letouzey J, Thiebaut B & Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. Genetics 157, 389–397.

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, Bâ K, Piry S, Lippens C, Diagne CA, Kane M, Sow A, Diallo M, Niang Y, Konecny A, Sarr N, Artige E, Charbonnel N, Granjon L, Duplantier JM & Brouat C (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.

Diallo A, Ndam NT, Moussiliou A, Dos Santos S, Ndonky A, Borderon M, Oliveau S, Lalou R & Le Hesran J-Y (2012) Asymptomatic Carriage of Plasmodium in Urban Dakar: The Risk of Malaria Should Not Be Underestimated. *PLoS ONE* **7**, e31100.

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.

Dray S, Said S & Debias F (2008) Spatial ordination of vegetation data using ageneralization of Wartenberg's multivariate spatial correlation. *Journal of Vegetation Science* **19**,45–56.

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.

Foll M & Gaggiotti O (2006) Identifying the Environmental Factors That Determine the Genetic Structure of Populations. *Genetics* **174**, 875–891.

Förster DW, Gündüz I, Nunes AC, Gabriel S, Ramalhinho MG, Mathias ML, Britton-Davidian J & Searle JB (2009) Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology* **18**, 4477-4494.

Gabriel SI, Stevens MI, da Luz MM & Searle JB (2012) Of mice and 'convicts': origin of the Australian house mouse, *Mus musculus*. *PloS ONE* **6**, e28622.

García-Rodríguez O, Andreou D, Herman JS, Mitsainas GP, Searle JB, Bonhomme F, Hadjisterkotis E, Schutkowski H, Stafford R, Stewart JR & Hardouin EA (2018) Cyprus as an ancient hub for house mice and humans. Journal of Biogeography 45, 2619-2630.

Gardner-Santana LC, Norris DE, Fornadel CM, Hinson ER, Klein SL & Glass GE (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.

Greenbaum G, Templeton AR, Zarmi Y, Bar-David S (2014) Allelic richness following population founding events – a stochastic modeling framework incorporating gene flow and genetic drift. PLOS ONE **9**:e115203. doi:10.1371/journal.pone.0115203. PMID:25526062.

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, Adams S, Jones EP, MacNicoll A, Mallon EB, Smith RH & Lambert MS (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, Xue AT, Alvarado-Serrano D, Boehm JT, Joseph T, Hickerson MJ & Munshi-South J (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.

Hothorn T, Buehlmann P, Dudoit S, Molinaro A & Van Der Laan M (2006) Survival Ensembles. *Biostatistics* **7**(3), 355--373.

Ihle S, Ravaoarimanana I, Thomas M & Tautz D (2006). An analysis of signatures of selective sweeps in natural populations of the house mouse. *Molecular Biology and Evolution* **23**, 790-797.

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.

Jassat W, Naicker N, Naidoo S & Mathee A (2013) Rodent control in urban communities in Johannesburg, South Africa: from research to action. *International Journal of Environmental Health Research* **23**, 474-483.

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

Jones EP, Johannesdottir F, Gündüz I, Richards MB & Searle JB (2011) The expansion of the house mouse into north-western Europe. *Journal of Zoology* **283**, 257–268.

Jones EP, Skirnisson K, McGovern TH, Gilbert MTP, Wilerslev E & Searle JB (2012) Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *Bmc Evolutionary Biology* **12**, 35.

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

Jones ZM & Linder F (2017) edarf: Exploratory Data Analysis using Random Forests. R package version 1.1.1. https://CRAN.R-project.org/package=edarf

Kajdacsi B, Costa F, Hyseni C, Porter F, Brown J, Rodrigues G, Farias H, Reis MG, Childs JE, Ko AI & Caccone A (2013) Urban population genetics of slum-dwelling rats (*Rattus norvegicus*) in Salvador, Brazil. *Molecular Ecology* **22**, 5056-5070.

King CM (2016) How genetics, history and geography limit potential explanations of invasions by house mice *Mus musculus* in New Zealand. *Biological Invasions* **18**, 1533-1550.

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, Estoup A, Hima K, Loiseau A, Tatard C, Dalecky A, Bâ K, Kane M, Diallo M, Sow A, Niang Y, Piry S, Berthier K, Leblois R, Duplantier JM & Brouat C (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, Granjon L, Dalecky A, Fossati O, Le Fur J, J.M. D & Handschumacher P (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.

MacKay JWB, Alexander A, Hauber ME, Murphy EC & Clout MN (2013) Does genetic variation among invasive house mice in New Zealand affect eradication success? *New Zealand Journal of Ecology* **37**, 18-25.

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.

Marston CG., Danson FM, Armitage RP, Giraudoux P, Pleydell D, Wang Q, Qui J & Craig PS (2014) A random forest approach for predicting the presence of Echinococcus multilocularis intermediate host Ochotona spp. presence in relation to landscape characteristics in western China. *Applied Geography* **55**, 176-183. DOI : 10.1016/j.apgeog.2014.09.001

https://prodinra.inra.fr/record/287109 Masi E, Pino FA, Santos MDS, Genehr L, Albuquerque JOM, Bancher AM & Alves JCM (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**, 247 –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.

Nei M, Maruyama T & Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1-10.

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York, USA. 512 p.

Ndiaye I. (2015) Étalement urbain et différenciation sociospatiale à Dakar (Sénégal). *Cahiers de géographie du Québec* **59** (166), 47–69. https://doi.org/10.7202/1034348ar

Parreira B & Chikhi L (2015) On some genetic consequences of social structure, mating systems, dispersal, and sampling. *Proceedings of the National Academy of Sciences* **112**, E3318-E3326.

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, Chapuis M, Gauffre B, Papaïx J, Cruaud A & Berthier K (2016) Mapping Averaged Pairwise Information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution* **7**, 1463-1475.

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, Park J, Combs M, Blum MJ, Bryant JE, Caccone A, Costa F, Deinum EE, Esther A, Himsworth CG, Keightley PD, Ko A, Lundkvist Å, McElhinney LM, Morand S, Robins J, Russell J, Strand TM, Suarez O, Yon L & Munshi-South J (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.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.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.

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

Richardson JL, Burak MK, Hernandez C, Shirvell JM, Mariani C, Carvalho-Pereira TSA, Pertile AC, JA P-M, Pedra G, Serrano S, Taylor J, Carvalho M, Rodrigues G, Costa F, Childs JE, Ko AI &

Caccone A (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.

Roberts DR & and Hamann A (2015) Glacial refugia and modern genetic diversity of 22 western North American tree species. Proceedins of the Royal Society, B 282. http://doi.org/10.1098/rspb.2014.2903

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

Rouhana S, Ranarifidy DN, Sinet A, Chomentowski V, Avner P & Godin LM (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, Ferrer-Mata A, Sánchez-DelBarrio JCC, Guirao-Rico S, Librado P, Ramos-Onsins SE & Sánchez-Gracia A (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.

Storz JF, Baze M, Waite JL, Hoffmann FG, Opazo JC & Hayes JP (2007) Complex signatures of selection and gene conversion in the duplicated globin genes of house mice. *Genetics* **177**, 481-500.

Strobl C & Boulesteix AL, Kneib T, Augustin T & Zeileis A (2008) Conditional Variable Importance for Random Forests. BMC bioinformatics, 9: 307

Strobl C, Hothorn T & Zeileis (2009) Party on ! The R Journal, 1:2, pages 14-17

Suzuki H, Nunome M, Koinoshita G, Aplin KP, Vogel P, Kryuskov AP, Lin M-L, Han S-H, Maryanto I, Tsuchiya K, Ikeda H, Shiroishi T, Yonekawa H & Moriwaki K (2013) Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity* **111**, 375-390.

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

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.

Wright MN & Ziegler A (2017) ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software, 77(1), 1-17.

Table 1. Characteristics of the land cover classes describing Dakar contemporary cityscape as determined by Borderon et al. (2014) using satellite images, UN habitat typology and census data on population wealth. Merging of the initial land cover classes based on spatial multivariate analysis and historical information is presented in the column "Final land cover classes"

Initial land cover classes	Characteristics	Final land cover classes
Old city blocks	Regular planned housing with concrete roads and no sand. Good equipment (water supply networks, sewers) and good accessibility. Mainly apartments and high-rise houses, old and commercial district. Predominance of tenants and significant presence of non-Africans. *Socio-economic status: mainly high*	**Old city**
Village-like	Anarchic habitat scattered along the peninsula, characterized by a high density of buildings, small plots, low-rise houses and sandy narrow roads. Typically, traditional fishing villages where inhabitants have always maintained customary law on the land. *Socio-economic status: very poor to medium high*	
Regular spontaneous	Approved self-building. Grid area with tarred roads between subdivisions and sand roads between land plots. Predominance of multi-storey, solid houses, often concrete. *Socio-economic status: very poor to medium high*	**Spontaneous**
Irregular spontaneous	Unplanned housing, often occurring without property title, resulting from important rural migration and moves from downtown during the rapid city sprawl in the 1970s and 1980s. Vehicle accessibility to these areas, often called "floating neighborhoods", is difficult due to narrow sandy streets and flooding. No equipment. *Socio-economic status: mainly very poor*	
High standing residential	Individual two-storey houses, villas with gardens or secured residences. High quality equipment and facilities. *Socio-economic status: mainly medium high*	**Residential**
Industrial areas	Mainly non-residential areas with large warehouses, factories, or official administrative buildings. *Socio-economic status: poor and medium*	**Industrial**
Vegetation	Grass, bush or tree cover mixed or not with various type of bare ground such as concrete or sand.	**Vegetation**

Table 2. 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 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	ρ	a_r	H_s	F_{IS}	n_{mt}	$H_{D\text{-}loop}$
Hann pêcheur	HPE	14.724035	-17.429230	47	0.043	4.60 ± 1.21	0.64 ± 0.15	0.16	3	H1
Ouakam	OKM	14.723088	-17.490068	46	0.108	4.65 ± 1.52	0.58 ± 0.18	0.20	4	H1
Ngor	NGR	14.748939	-17.517352	31	0.129	4.42 ± 1.13	0.57 ± 0.16	0.23	4	H1
Park of Hann	PAH	14.726967	-17.434613	19	-0.021	4.67 ± 1.68	0.64 ± 0.17	0.14	–	–
Yoff	YOF	14.757200	-17.473321	31	0.054	4.86 ± 1.86	0.60 ± 0.22	0.23	4	H1
Plateau-Rebeuss	PRB	14.672396	-17.442259	54	0.030	5.71 ± 1.55	0.64 ± 0.13	0.26	–	–
Plateau- Faidherbe	PFD	14.674027	-17.436326	39	0.032	5.52 ± 1.73	0.66 ± 0.14	0.11	4	H2/H16
Port of Dakar	POR	14.683414	-17.430436	29	0.001	6.11 ± 2.21	0.67 ± 0.16	0.16	8	H1/H16
Grand Dakar	GDK	14.704387	-17.455356	34	0.056	5.09 ± 1.44	0.59 ± 0.18	0.24	–	–
Patte d'Oie	PDO	14.748332	-17.437953	39	0.036	5.08 ± 1.77	0.68 ± 0.11	0.30	–	–

Colobane	CBN	14.695814	-17.444924	45	0.040	4.84 ± 1.46	0.65 ± 0.12	0.16	–	–
Gueule Tapée	GTP	14.680193	-17.458608	43	0.046	4.86 ± 1.36	0.61 ± 0.15	0.30	4	H1/H2/H17
Gorée Island	IDG	14.667140	-17.398320	72	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	0.077	4.77 ± 1.61	0.58 ± 0.19	0.31	5	H1/H2

Table 3 – Results of the CAR model according to the combination of covariates considered. For each variable, the first row indicates the posterior probability (i.e. considering a threshold of 0.05, a posterior probability > 0.95 indicates a significant positive effect while a posterior probability < 0.05 indicates a significant negative effect of the variable). The regression coefficient with its standard deviation and 95%CI is indicated in italics for each variable. For each model, the table shows: the correlation coefficient of Pearson (*r*) between predicted and observed values, the p-value of the Kolgomorov-Smirnov test (*ks*) on the PIT values and the wAIC and DIC values.

	Cityscape variables							**Fit**	
Model	**Old_City**	**Spontaneous**	**Residential**	**Industrial**	**Vegetation**	**Built-up**	**Connection**	*r* *ks*	**wAIC** **DIC**
	0.165	0.627	1.000	0.000	1.000				
Model 1	*-0.010±0.010*	*0.002±0.008*	*0.048±0.012*	*-0.100±0.02*	*0.049±0.008*			0.95	-1418
	[-0.029 ; 0.010]	*[-0.014 ; 0.019]*	*[0.024 ; 0.071]*	*[-0.148 ;-0.052]*	*[0.034 ; 0.064]*			0.509	-1475
	0.179	0.553	1.000	0.000	1.000	0.297			
Model 2	*-0.009±0.010*	*0.001±0.009*	*0.048±0.012*	*-0.099±0.024*	*0.048±0.008*	*-0.003±0.006*		0.95	-1421
	[-0.029 ; 0.011]	*[-0.016 ; 0.018]*	*[0.024 ; 0.072]*	*[-0.147 ;-0.052]*	*[0.033 ; 0.063]*	*[-0.015 ; 0.009]*		0.536	-1479
	0.943	0.003	1.000	0.001	1.000		0.000		
Model 3	*0.014±0.009*	*-0.021±0.007*	*0.061±0.010*	*-0.066±0.021*	*0.027±0.007*		*-0.072±0.007*	0.97	-1509
	[-0.003 ; 0.031]	*[-0.036 ;-0.006]*	*[0.040 ; 0.081]*	*[-0.107 ;-0.026]*	*[0.013 ; 0.040]*		*[-0.086 ;-0.058]*	0.420	-1562
	0.932	0.013	1.000	0.001	1.000	1.000	0.000		
Model 4	*0.013±0.008*	*-0.017±0.007*	*0.059±0.010*	*-0.066±0.020*	*0.028±0.007*	*0.017±0.005*	*-0.081±0.007*	0.99	-1519
	[-0.004 ; 0.030]	*[-0.032 ;-0.003]*	*[0.039 ; 0.079]*	*[-0.106 ;-0.026]*	*[0.015 ; 0.041]*	*[0.006 ; 0.027]*	*[-0.096 ;-0.066]*	0.514	-1571

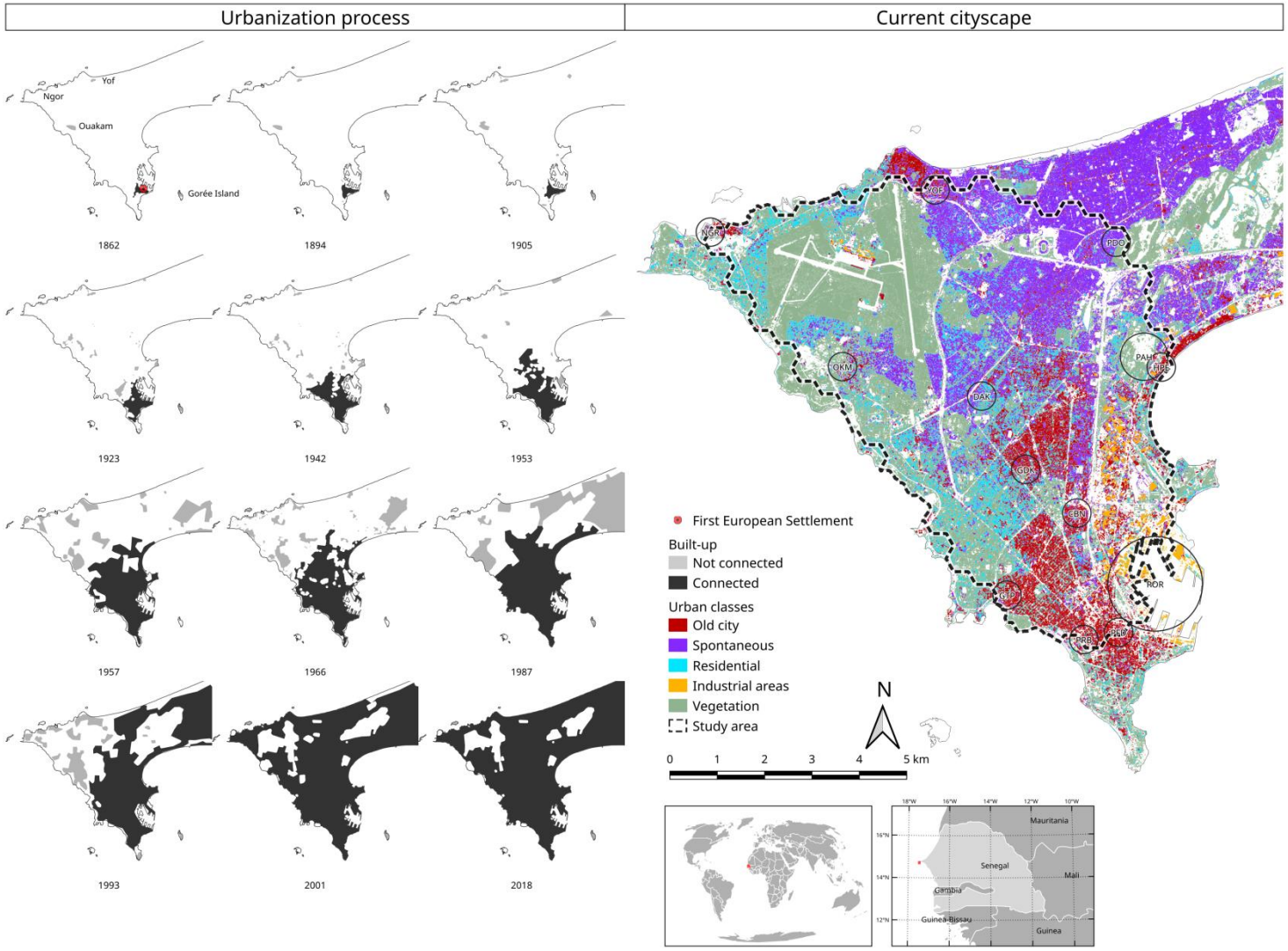
Figure Legends

Figure 1 – Historical and current Dakar cityscape. Historical cityscape is represented by a ime series of 12 maps depicting the spatial expansion of Dakar between 1862 and 2018. The location of the first European settlement is represented with a red dot on the 1862 map while the grey polygons represent build-up areas (dark or light grey depending on whether they are connected to the first European Settlement). For each map the coastline correspond to the current one. The localization of the old fishermen villages of Ouakam, Yoff and Ngor are indicated on the first historical map. Current cityscape of Dakar based on historical information and results of the spatial PCA performed on the 100x100m resolution raster of the typology of socioeconomic habitats retrieved from Borderon *et al*. (2014) – see Supplementary material, section 1 for details. The study area for the house mouse is delineated with a dashed line.

Figure 2 - Spatial genetic structure of *Mus musculus domesticus* in Dakar assessed using: (A) the five mitochondrial haplotypes found in Dakar – The median joining network is presented with white squares corresponding to network nodes, and blue, orange and yellow symbols corresponding to haplotypes from the three haplogroups (HG4, HG1 and HG11, respectively) identified by Bonhomme *et al.* (2011) ; (B) 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; (C) GESTE with a color scale representing the variation in local-F_{ST} from lowest values in red to highest values in blue ; (D) 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

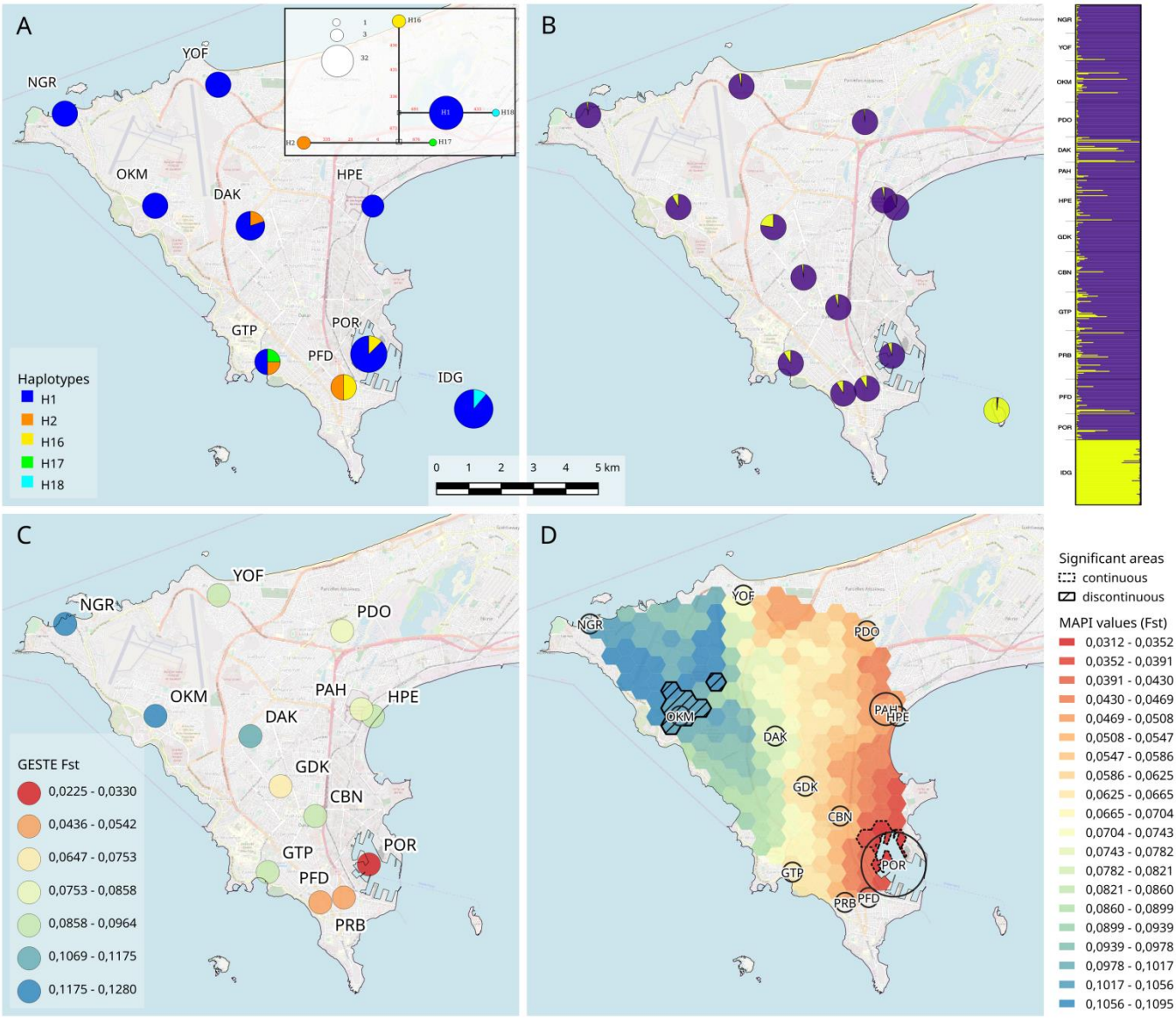
sampling sites are represented by a circle with a size depending on the spatial distribution of individuals sampled within each site. Small inserts representing the location of Senegal and of Dakar were generated using data from Natural Earth (public domain).

Figure 1



Urbanization process
Current cityscape
Yof
Ngor
Ouakam
Gorée Island
1862
1894
1905
1923
1942
1953
1957
1966
1987
1993
2001
2018
YOF
NGR
PDO
OKM
PAH
HPE
DAK
GDK
CBN
ROR
GTP
PRB
PRD
First European Settlement
Built-up
Not connected
Connected
Urban classes
Old city
Spontaneous
Residential
Industrial areas
Vegetation
Study area
N
0
1
2
3
4
5 km
Mauritania
Senegal
Mali
Gambia
Guinea-Bissau
Guinea

Figure 2


A
NGR
YOF
OKM
DAK
HPE
GTP
POR
PFD
IDG
Haplotypes
H1
H2
H16
H17
H18
0 1 2 3 4 5 km
B
C
GESTE Fst
0,0225 - 0,0330
0,0436 - 0,0542
0,0647 - 0,0753
0,0753 - 0,0858
0,0858 - 0,0964
0,1069 - 0,1175
0,1175 - 0,1280
PDO
PAH
GDK
CBN
PRB
D
Significant areas
continuous
discontinuous
MAPI values (Fst)
0,0312 - 0,0352
0,0352 - 0,0391
0,0391 - 0,0430
0,0430 - 0,0469
0,0469 - 0,0508
0,0508 - 0,0547
0,0547 - 0,0586
0,0586 - 0,0625
0,0625 - 0,0665
0,0665 - 0,0704
0,0704 - 0,0743
0,0743 - 0,0782
0,0782 - 0,0821
0,0821 - 0,0860
0,0860 - 0,0899
0,0899 - 0,0939
0,0939 - 0,0978
0,0978 - 0,1017
0,1017 - 0,1056
0,1056 - 0,1095