Received: 20 August 2019 | Accepted: 8 April 2020

DOI: 10.1111/mec.15451

FROM THE COVER

MOLECULAR ECOLOGY WILEY

Genes acting in synapses and neuron projections are early targets of selection during urban colonization

Jakob C. Mueller[1] | Martina Carrete[2,3] | Stefan Boerno[4] | Heiner Kuhl[4,5] | José L. Tella[2] | Bart Kempenaers[1]

[1]Department of Behavioural Ecology & Evolutionary Genetics, Max Planck Institute for Ornithology, Seewiesen, Germany

[2]Department of Conservation Biology, Estación Biológica de Doñana – CSIC, Sevilla, Spain

[3]Department of Physical, Chemical and Natural Systems, University Pablo de Olavide, Sevilla, Spain

[4]Sequencing Core Facility, Max Planck Institute for Molecular Genetics, Berlin, Germany

[5]Department of Ecophysiology and Aquaculture, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany

Correspondence
Jakob C. Mueller, Department of Behavioural Ecology & Evolutionary Genetics, Max Planck Institute for Ornithology, Seewiesen, Germany.
Email: mueller@orn.mpg.de

Funding information
MINECO, Spain, Grant/Award Number: CGL2012-31888 and and CGL2015-71378-P; Fundación Repsol; Severo Ochoa microproyecto award from Estación Biológica de Doñana; Max Planck Society

Abstract

When a species colonizes an urban habitat, differences in the environment can create novel selection pressures. Successful colonization will further lead to demographic perturbations and genetic drift, which can interfere with selection. Here, we test for consistent urban selection signals in multiple populations of the burrowing owl (*Athene cunicularia*), a species that colonized South American cities just a few decades ago. We sequenced 213 owls from three urban-rural population pairs and performed a genome-wide comparison of urban against rural birds. We further studied genome-wide associations with flight initiation distance, a measure of harm avoidance in which urban and rural birds are known to differ. Based on four samples taken over nine years from one of the urban populations, we investigated temporal allele frequency changes. The genomic data were also used to identify urban-specific signatures of selective sweeps. Single genomic sites did not reach genome-wide significance for any association. However, a gene-set analysis on the strongest signals from these four selection scans suggests a significant enrichment of genes with known functions related to synapses and neuron projections. We identified 98 genes predominantly expressed in the brain, of which many may play a role in the modulation of brain connectivity and consequently in cognitive function and motivational behaviour during urbanization. Furthermore, polymorphisms in the promoter region of the synaptic *SERT* gene – one of the two candidates known to correlate with urban colonization in birds – associated with the habitat in which individuals lived (urban vs. rural).

KEYWORDS

functional enrichment, gene set analysis, GWAS, neuron projection, selection signature, Strigiformes, synapse

1 | INTRODUCTION

The extent of urban habitats is increasing worldwide, and most humans now live in cities (United Nations, 2018). Urbanization results in dramatic environmental changes, including habitat fragmentation, increased temperature, altered availability of resources and breeding sites, reduced predation or parasite pressure and air, noise, and light pollution (Johnson & Munshi-South, 2017). Despite the

associated general decline in diversity, there are many examples of species that successfully colonize urban habitats. During and shortly after such successful colonization events, strong episodes of selection are expected. Among birds, urban colonization has been related to intraspecific variation in a wide range of behavioural traits, including antipredator and dispersal behaviour (Alberti et al., 2016; Evans, Hatchwell, Parnell, & Gaston, 2010). Urban life may affect daily and seasonal activity patterns like the timing of dawn singing (Kempenaers, Borgstrom, Loes, Schlicht, & Valcu, 2010), reproduction (Chamberlain et al., 2009), and migratory and dispersal behaviour (Evans et al., 2012). Increased sedentariness and a shifted breeding period may then restrict gene flow between urban and rural populations, thus facilitating local urban adaptation (Evans et al., 2012). It has also been proposed that urban life selects for individuals that show a reduced reactivity to human disturbance in a number of bird species (Carrete & Tella, 2011).

Colonization of urban habitats by animals and plants is ideal to study recent adaptive and nonadaptive processes at the genome level. The habitats of city centres and suburban areas and their historical development appear to be similar across a wide geographical range and thus represent repeated natural experiments to study evolution in action (Donihue & Lambert, 2015; Johnson & Munshi-South, 2017; Marzluff, 2016). Urban colonization typically goes hand-in-hand with demographic perturbations such as population bottlenecks or fragmentation, which can lead to similar genome signatures as adaptive processes, although across all loci (Munshi-South, Zolnik, & Harris, 2016). Studying repeated urban settlements by the same species can be useful to disentangle random genome-wide demographic effects from local deterministic effects based on selection. While genetic drift produces random allele frequency shifts at all sites, similar selective pressures among different cities could lead to the same selected regions or regulatory pathways in the genome. It is thus important to analyse multiple cities, and ideally those for which the genomic population structure and demographic history is known.

In animals, behaviour is among the most likely traits under early selection in response to urban colonisation (Evans et al., 2010). Behaviour determines where an individual lives, where it nests, where it obtains resources, how it responds to threats (e.g., how it avoids predators and deals with competitors), how it chooses a mate, and so on. Changes in behavioural traits may expose individuals to new environments and, vice versa, new environments may promote changes in behaviour (Zuk, Bastiaans, Langkilde, & Swanger, 2014). The study of genetic correlates of behavioural variation in vertebrates is still in its infancy, but prominent candidates can be determined based on the neuronal and hormonal signalling systems and its regulating factors (Adkins-Regan, 1996; Elphick, Mirabeau, & Larhammar, 2018; Yap & Greenberg, 2018). In particular, genes that code for synaptic proteins have attracted attention in the context of persistent behavioural or personality variation and in the context of colonization events. For example, two European invasive populations of the yellow-crowned bishop (*Euplectes afer*) showed allele frequency differences at the synaptic dopamine receptor *D4* gene (*DRD4*) in comparison to the native African population, which were consistent with allele frequency changes observed during early introduction stages (Mueller et al., 2017). Similarly, variation at the *DRD4* locus was associated with wariness and urban site selection in black swans (*Cygnus atratus*) (van Dongen, Robinson, Weston, Mulder, & Guay, 2015). A microsatellite in the synaptic serotonin transporter gene (*SERT*), a candidate gene for harm avoidance behaviour, exhibited an association with urban habitat across 10 of 12 tested urban-rural population pairs of the blackbird (*Turdus merula*) (Garroway & Sheldon, 2013; Mueller, Partecke, Hatchwell, Gaston, & Evans, 2013).

Here, we analyse the genomic signals of adaptation in populations of burrowing owls (*Athene cunicularia*) that recently colonized Argentinian cities in comparison to neighbouring rural populations. The burrowing owl is a species typically associated with open grasslands in North and South America, that has begun breeding in South American suburban areas within the last few decades (Carrete & Tella, 2011). Historically, burrowing owls associated with fossorial mammals such as plains viscachas (*Lagostomus maximus*), whose burrows they used for nesting (Martinez, Baladron, Cavalli, Bo, & Isacch, 2017). However, the owls of the Southern range evolved the ability to excavate their own burrows (Martinez et al., 2017), which may have facilitated moving into the urban environment (Rebolo-Ifrán, Tella, & Carrete, 2017). Our genomic data indicate that each city was independently colonized and that restricted gene flow occurred between neighbouring urban and rural populations, but there is no gene flow between populations from different cities (Mueller et al., 2018). Estimated effective population sizes in the recent past were consistently lower for the urban populations compared to the rural ones (Mueller et al., 2018). Despite lower population size, urban populations appear to be quite successful: they show higher breeding densities than rural populations (Rodríguez-Martínez, Carrete, Roques, Rebolo-Ifrán, & Tella, 2014), presumably linked to reduced predation pressure (Rebolo-Ifrán et al., 2017). Moreover, changes in behavioural traits due to selection or selective immigration have been documented in urban populations. Urban birds showed on average a smaller flight initiation distance (FID) upon approach than rural ones (Carrete & Tella, 2011; Rebolo-Ifrán et al., 2015). FID – a measure of risk perception – is an individually-consistent and heritable trait (Carrete et al., 2016; Carrete & Tella, 2017). This makes the burrowing owl a suitable system to study rapid adaptation to an urban environment.

In this study, we analysed full genome sequences of 213 individuals (mapped against the recently assembled and annotated burrowing owl genome [Mueller et al., 2018]) from three urban populations, three associated rural populations and four temporal samples taken across nine years in one of the urban populations. We applied four independent tests for selection: (a) consistent allele frequency shifts between urban-rural population pairs; (b) allele frequency changes over time in one urban population; (c) genomic selection signals ("selective sweeps") in the urban populations; and (d) genome-wide associations with FID. We used these tests to address the following specific questions. (a) Are there any signals at single genomic

sites of genome-wide significant selection? Given the high number of tested single nucleotide polymorphisms (SNPs), only sites under strong selection can be detected; (b) are there functionally-related gene groups in the top genomic regions of strongest selection signals? This reveals also weak selection in functionally-related genes and is considered a powerful complementary genome-wide approach in addition to the search for genome-wide significant single site signals (Fridley & Biernacka, 2011; Mooney, Nigg, McWeeney, & Wilmost, 2014; Wang, Jia, Wolfinger, Chen, & Zhao, 2011); (c) are there consistent selection signals in the urban population (Bahia Blanca) where all four tests were performed? and (d) do the candidate genes *SERT* and *DRD4* show signatures of selection?

2 | MATERIALS AND METHODS

2.1 | Sampling

As part of a long-term study, urban and rural burrowing owls living in and around Bahia Blanca city, Argentina had been individually marked, monitored and blood-sampled since 2006 (Carrete & Tella, 2011; Rodríguez-Martínez et al., 2014). Between 2012 and 2016, breeding burrowing owls had also been sampled in, and near, two additional Argentinian urban settlements of the pampas, Sierra de la Ventana and Tandil, located 77 and 314 km linear distance from Bahia Blanca, respectively (see figure 1 in Mueller et al., 2018). All blood samples were stored in ethanol.

Breeding habitats of the owls were classified as either rural or urban following the criteria described in Carrete and Tella (2011) and Rebolo-Ifrán et al. (2015). "Urban nests" were defined as those located in private and public gardens and along streets, usually within 10–100 m of inhabited buildings. "Rural nests" were those found outside the city border in grasslands and pastures with wide-ranging livestock where human presence and activities were minimal.

For our study, we defined a priori seven sampling sites for the three urban-rural comparisons (figure 1 in Mueller et al. [2018], Table S1 and Data S1): Bahia Blanca urban 1 and 2 ($BB1_{urban,}$ $BB2_{urban}$), Bahia Blanca rural (BB_{rural}), Sierra de la Ventana urban (SV_{urban}) and rural (SV_{rural}), Tandil urban (TA_{urban}) and rural (TA_{rural}). From each of these sampling sites, we randomly selected 20 breeding individuals from 2012 to 2016 (except for SV_{rural} where we only sampled 17 individuals), excluding first-degree relatives based on ringing data of parents and their offspring. For the largest and more intensively monitored city (Bahia Blanca) we sampled 20 individuals from each of two spatially separated suburban areas ($BB1_{urban}$ and $BB2_{urban}$) to test the robustness of selective sweep inference based on simulations of 20 individuals (see below). For $BB1_{urban}$, we also sampled individuals over a nine-year period (2006: $N = 17$, 2009: $N = 20$, 2012: $N = 20$ and 2015: $N = 20$). We further included 19 individuals from the urban-rural transition zone close to $BB2_{urban}$ for which FID had been measured ($BB_{transition}$). For details of the FID measurement see Carrete and Tella (2013). Thus, in total, we used blood samples from 213 individuals for genomic analysis.

2.2 | Sequencing, reference mapping and SNP calling

We extracted DNA from all blood samples with the Blood QuickPure kit (Macherey-Nagel) applying a predigestion with Proteinase K in Digsol buffer. All 213 birds were individually sequenced with the Illumina HiSeq 2500 technique using a 200–300 bp insert paired-end library with 125 bp read length at the Sequencing Core Facility of the MPI for Molecular Genetics in Berlin.

For each individual, we mapped reads against the reference genome using Bowtie2 (Langmead & Salzberg, 2012). We called variants using the GATK HaplotypeCaller in gvcf mode and assigned individual genotypes by a joint genotype calling of all samples (Van der Auwera et al., 2013). The SNPs were then quality-filtered following fixed rules from the GATK best practices recommendations (Van der Auwera et al., 2013). We excluded rare SNPs with minor allele count of less than three and low-quality SNPs with 20 or more missing genotypes across the 213 individuals. Additionally, to avoid SNP calls from misassembled paralogous genome regions, we only included SNPs if read depth was smaller than the mean coverage plus five standard deviations across all SNPs. For subsequent analyses, we also excluded Z-chromosomal SNPs outside the pseudoautosomal region (PAR) to avoid the different ploidy in males and females. The PAR of the Z chromosome was defined on scaffold chrZ_part1: 34225000–36080512 according to the differential sex patterns of read coverage, heterozygosity and Hardy-Weinberg deviations. The final data set of 213 individuals contained 11,699,304 SNPs (on average 1 SNP per ~100 bp), of which 4,199,164 SNPs were common with a minor allele frequency >0.05. The three association analyses (see below) are based on these common SNPs only.

2.3 | Data analysis

First, we investigated overall signals of selection during urbanization by testing the association between genotype (additive allele effect model) and habitat (urban vs. rural) for each common SNP combining data from the three cities. We applied a mixed effects logistic regression model using the R package GENESIS (Gogarten et al. 2019). The 137 individuals of $BB1_{urban}$ from 2012, $BB2_{urban}$, BB_{rural}, SV_{urban}, SV_{rural}, TA_{urban} and TA_{rural} were adjusted for background genetic structure by including a genetic relationship (relatedness) matrix (grm) as random effect. The grm matrix was calculated from 1,493,014 independent SNPs (composite genotypic linkage disequilibrium <0.5) according to Yang, Lee, Goddard, and Visscher (2011) using the R package SNPRELATE (Zheng et al., 2012). The distribution of genetic relationship values is shown in Figure S1. The model design prevented inflation of test results, as seen by the close fit between observed and expected uniform distributed *p*-values in Figure S2a.

Second, we investigated evidence for recent and consistent (between 2006 and 2015) selection by testing for an association between genotype (additive allele effect model) and sampling year (2006, 2009, 2012, 2015) for each common SNP using the R package

GENESIS with a linear mixed model. The 77 individuals of the temporal samples of $BB1_{urban}$ were adjusted for background genetic structure by including the grm matrix (see above) as random effect. The model design prevented inflation of test results (Figure S2b).

Third, we investigated the association between genotypes of common SNPs and FID. Using the 132 individuals for which FID had been measured ($BB1_{urban}$ – including the temporal samples -, $BB2_{urban}$, BB_{rural}, SV_{urban} and the extra birds from the transition zone of $BB_{transition}$), we applied a linear mixed model with the grm matrix (see above) as random effect to adjust for background genetic structure and sex as fixed factor to adjust for known sex differences in FID (Carrete & Tella, 2013). Again, the model design sufficiently prevented inflation of test results (Figure S2c).

For each of the three association tests described above, we summarized single-SNP results in 86,232 half-overlapping 25 kb windows (25 kb equals about the average gene size and should be appropriate when the downstream analyses are gene-based) by calculating the sum of the nominal significant ($p < .05$) −log(*p*-values). This parameter – here named "Σp" – was designed to measure both the accumulation and the strength of association signals, similar to parameters used in gene-set analyses (de Leeuw, Mooij, Heskes, & Posthuma, 2015; Mooney & Wilmot, 2015). To evaluate genome-wide significant thresholds, we performed the three tests 100 times, but with permutated response variables. We retained potential phenotypic covariance structure of the quantitative response variable FID by permuting only the residuals according to Abney (2015) using the R package MVNPERMUTE. The genome-wide significant parameter threshold was defined as the level above which only five of the 100 simulations had at least one window with higher values.

Fourth, we investigated selective sweeps specific for urban populations using diploS/HIC (Kern & Schrider, 2018). A supervised machine-learning algorithm on genomic simulations with neutral, soft, linked soft, hard and linked hard sweep evolution (2,000 simulations of 275 kb for each scenario) was used to classify (with probabilities) all neighbouring 25 kb windows (41,608 windows) in the populations $BB1_{urban}$ from 2012, $BB2_{urban}$, BB_{rural}, SV_{urban}, SV_{rural}, TA_{urban}, TA_{rural} into hard sweep, soft sweep, neighbouring hard sweep, neighbouring soft sweep and neutral windows. The genomic coalescent simulations were based on simplified demographics of a characteristic urban and rural population (Figure S3, derived from figure 4 of Mueller et al. [2018]) using discoal (Kern & Schrider, 2016). Complete hard and soft sweeps were simulated with selection coefficients between 0.01 and 0.001 and a starting minor allele frequency of 0.02 in the case of the soft sweep. The learning algorithm employs 12 major summary statistics (normalized values) of selective sweeps including diversity measures, allele frequency spectra, linkage disequilibria patterns and haplotype structure in the focal and neighbouring windows (Kern & Schrider, 2018). As suggested for weak population-scaled selection (as in urban populations), we combined soft and hard sweep probabilities and calculated urban-specific sweep probabilities as urban sweep probability (soft and hard) minus rural sweep probability (soft and hard) for all urban-rural comparisons (mean($BB1_{urban}$, $BB2_{urban}$)-BB_{rural}, SV_{urban}-SV_{rural}, TA_{urban}-TA_{rural}). The overall urban-specific sweep probability is then the average across the three urban-rural comparisons (here named "P_{sweep}"). We arbitrarily considered a probability larger than 0.8 as important.

For the Bahia Blanca samples, where the sample scheme allowed all four independent tests, we also summarized the four tests into a single statistic. First, we performed all four tests for the Bahia Blanca samples only and then combined the results by summing up the ranks of each test for each genomic window ("Σrank").

Complementing the scans for genome-wide significant single window signals, we performed tests for enrichment of functionally-related gene sets in the windows with highest Σp, P_{sweep} or lowest Σrank parameters using the gene ontology (GO) annotations. It has been recommended to use the top 10–400 genes to have sufficient power for the detection of the differently sized GO terms (Mooney & Wilmot, 2015; Ramanan, Shen, Moore, & Saykin, 2012). We decided to test the top 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1,000 windows, which associated approximately to between 10 and 400 genes. This scheme, suitable for overlapping windows of the association tests, translates to the top 25, 50, 100, 150, 200, 250, 300, 350, 400, 450 and 500 windows for the selective sweep test and for the combined analysis of the Bahia Blanca sample, where nonoverlapping windows were evaluated. We performed standard candidate-vs.-background enrichment analyses using hypergeometric tests implemented in the R package GOFUNCR (Grote, 2018). We used the NCBI gene annotation of our genome assembly (NCBI *Athene cunicularia* Annotation Release 100), which contains 13,166 unique protein-coding genes with known gene symbols in all autosomes and the Z-chromosomal PAR region, to define the set of all background genes and to define candidate genes as those falling within the selected windows (i.e, within a maximum distance of 5 kb). We corrected for testing multiple GO terms within each of the three main GO categories/families "biological process", "cellular component" and "molecular function" by evaluating 1,000 simulated data sets with random candidate regions. The family-wise error rate (FWER) is defined as the fraction of random sets where the lowest *p*-value across all GO terms is lower or equal to the raw *p*-value of the tested GO term. This procedure of permuting candidate regions of fixed size within a fixed annotated background (all chromosomes) takes into account that genes of different lengths differ in their false positive association probability. Of note, this genome-wide permutation corrects for multiple testing and avoids false positives in random sets of candidate regions. It is therefore not necessary to select a *priori* genome-wide significant candidate windows for this gene-set analysis. All significant GO terms with associated genes are plotted in a network using the function cnetplot of the R package CLUSTERPROFILER (Yu, Wang, Han, & He, 2012).

3 | RESULTS

3.1 | Genome-wide significant signals of single sites

Figure 1a–c show the results across all genomic windows of all three association tests for signals of selection. We did not find single

genome-wide significant windows that differentiated the urban and rural populations (Figure 1a), showed a temporal change (Figure 1b) or were associated with FID (Figure 1c).

We identified 49 windows with a mean urban-specific sweep probability larger than 0.8 (arbitrary threshold; see Section 2) (Figure 1d). These windows are associated with 38 genes with known gene symbols (Table S2). This gene list is not significantly enriched for any GO term. For comparison, there are only 14 windows showing a mean rural-specific sweep probability larger than 0.8.

3.2 | Gene enrichment analyses for the sites showing the strongest association signals

Assuming coselection of functionally related genes or pathways, we performed gene ontology enrichment analyses for the windows showing the strongest association signals (up to top 1,000 windows, see Section 2). This genome-wide corrected alternative approach identifies groups of functionally related genes in the windows showing the strongest association signals and thus complements the single-window analysis, which detects only strong signals depending on genome size. Following this approach, Figure 2 and Table S3 show all family-wise significant GO terms of the four selection tests and the combined test on the Bahia Blanca samples. We found a significant enrichment for many genes associated with the term "Synapse" for the urban-rural habitat association analysis (FWER for the top 1,000 windows: 0.001; Table S3). The top 1,000 windows contained 373 genes of which 60 were associated with the term "Synapse". The gene sets of all other significant terms related to synaptic functions ("Synapse part", "Postsynapse", "Postsynaptic membrane") were subsets of these 60 genes. A second related and overlapping enriched group of genes related to the terms "Intrinsic/Integral component of membrane" (FWER for the top 200 windows: 0.030; Figure 2, Table S3).

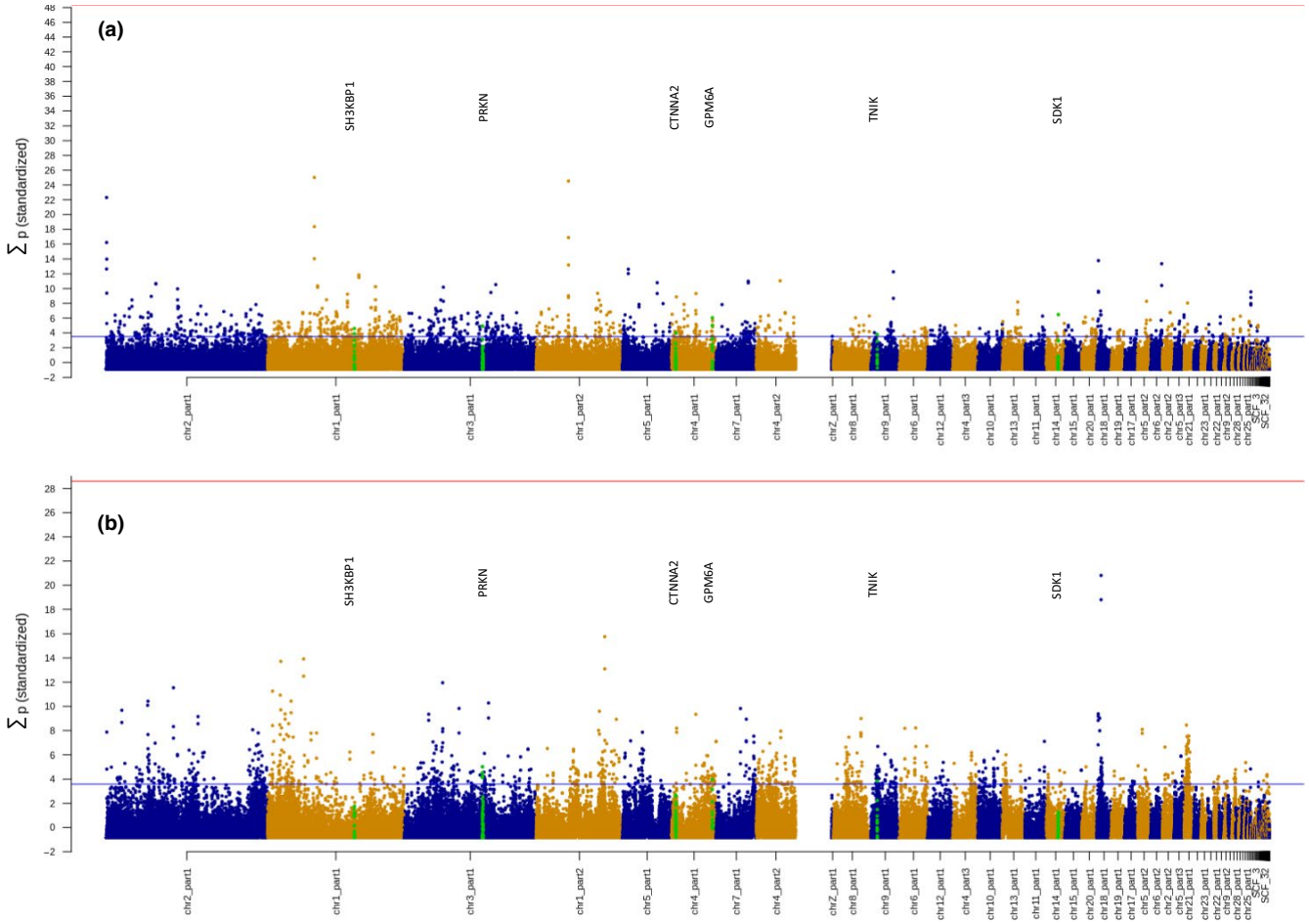

FIGURE 1 Manhattan plots showing for each genomic window (a) urban-rural habitat associations (combined analysis for all three urban-rural pairs), (b) temporal allele frequency changes in $BB1_{urban}$, (c) FID associations and (d) mean urban-specific selective sweep probabilities. For (a), (b) and (c) the genome-wide significant level is indicated by the red line and the top 1,000 windows used for the gene-set analyses are those above the blue line. For (d) the 0.8 probability thresholds of urban-specific (positive values) and rural-specific (negative values) selective sweeps are indicated by red lines and the top 500 windows are those above the blue line. The regions of the six shared genes from the enriched gene sets related to synapses and neuron projections of the urban-rural association and combined Bahia Blanca analyses (see Figure 2) are indicated in green and labelled. Scaffolds appear in alternating colours and scaffolds smaller than 7 Mb are labelled where possible

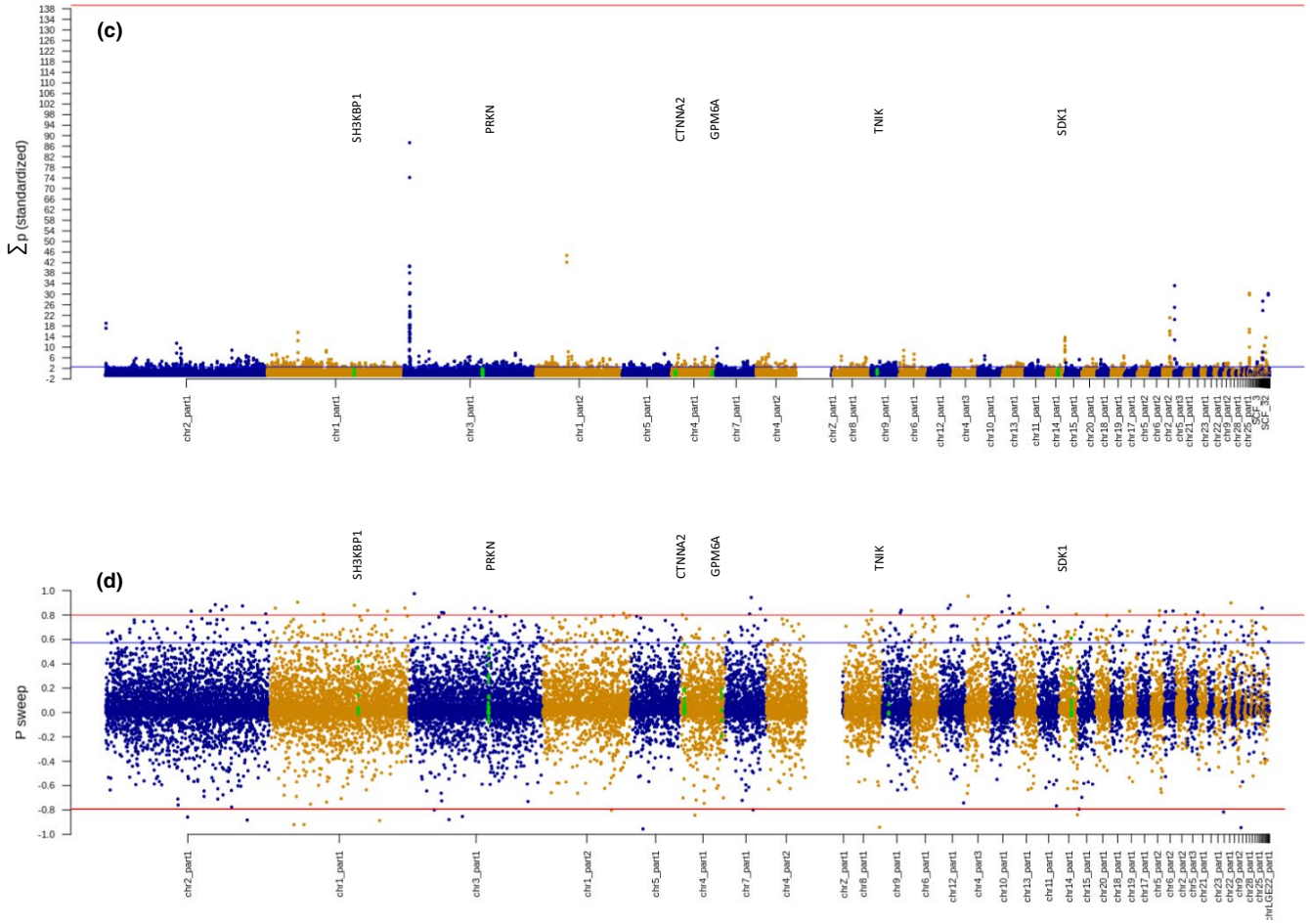

FIGURE 1 (Continued)

The top 500 windows for the temporal change associations showed significant enrichment of genes related to the term "Beta-tubulin binding" (FWER = 0.040). There was no significant enrichment for the top windows of the FID association and selective sweep analyses (all FWER > 0.05).

The ranks of the windows based on the three independent association tests (urban-rural habitat, temporal change and FID) in Bahia Blanca were all positively correlated (Table S4). The rank correlations with the windows of the selective sweep analysis were weakly negative when significant (Table S4). The top 500 windows of the combined test on the Bahia Blanca samples were significantly enriched for the term "Neuron projection development" with 44 genes (FWER = 0.011; Table S3). The genes of the other significantly enriched terms were either complete subgroups ("Neuron projection morphogenesis") or highly concordant groups to "Neuron projection development" with more than 93% overlap ("Cell part morphogenesis" and "Neuron development") (Table S3, Figure 2). Six of the genes of this functional cluster (*SH3KBP1*, *TNIK*, *PRKN*, *CTNNA2*, *GPM6A*, *SDK1*) co-occurred in the 60 enriched synaptic genes of the urban-rural habitat association analysis (Figure 2).

In summary, we found 98 genes acting in synapses or neuron projections among the 131 genes of all the enriched functional GO terms. Importantly, the genes of these enriched GO terms are mostly tagged by independent windows. Only two of the 156 associated windows tag two or three overlapping genes. Therefore, the enrichment results are not due to the clustering of genes in single windows.

3.3 | Candidate genes *SERT* and *DRD4*

The *DRD4* and *SERT* genes each span two nonoverlapping 25 kb windows and three and four overlapping ones, respectively. The standardized test parameters (Σp, P_{sweep}) for the *DRD4* windows did not reach nominal significance for being an outlier for strong selection (all values <1.645, i.e. the one-sided 5% significance threshold) in the four tests of selection and the *DRD4* windows were also ranked low in the combined test of the Bahia Blanca samples (rank 29641 and 29467 of 41608 windows, i.e. top 71.2% and 70.8%). However, the parameters of the first *SERT* window (chr19_part1: 4825000–4850000) were nominally significant for the urban-rural habitat association analysis (standardized value of Σp: 1.75 (p = .040) and relatively high for the FID association analysis (standardized value of Σp: 1.48). Also, the two windows spanning the whole *SERT* gene (chr19_part1: 4837500–4887500) ranked relatively high in the combined test of Bahia Blanca samples (rank 13053 and 2725 of 41608 windows, i.e. top 31.4% and 6.5%).

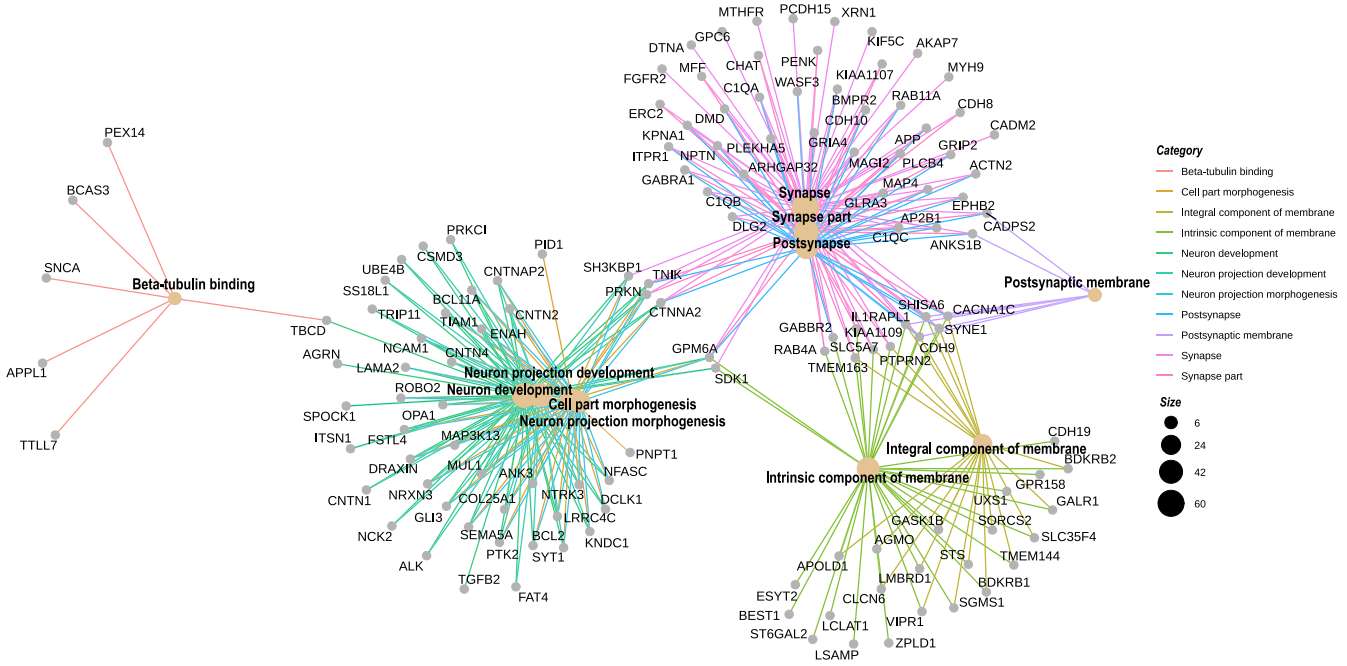

FIGURE 2 All significantly enriched GO terms and genes associated with the top 50–1,000 windows of the three association tests applied on overlapping windows (urban-rural, temporal and FID associations) and the top 25–500 windows of the selective sweep test and the combined test for the Bahia Blanca samples are shown in a network (cnetplot). Clearly, there is enrichment of many genes associated with synaptic and neuron projection functions. For details see Table S3

4 | DISCUSSION

Cities are growing worldwide and represent novel habitats to which wildlife can adapt. Repeated establishment of populations in urban areas allows the investigation of general urban adaptation, but also presents a challenge because of the demographic perturbations during colonisations. So far, only a few genomic scans for urban adaptation have been published, mostly using representative markers of the genome such as restriction-site associated DNA sequencing or transcriptome data (DeCandia et al., 2019; Harris & Munshi-South, 2017; Perrier et al., 2018; Ravinet et al., 2018; Theodorou et al., 2018). Here, we show how a young urbanization such as that of the burrowing owl in South America can be used to analyse adaptation to a new environment by complete genome resequencing of 213 individuals. We combined evidence from alternative analysis approaches based on multiple independent replicate samples. Given the recent history of urbanization in the burrowing owl, selection probably acted on standing genetic variation. We thus applied methods based on allele frequency changes between urban and rural habitats, over time in urban samples and related to a specific urban-related phenotype (flight-initiation distance, FID), and included soft sweeps in the sweep analysis. Single genomic windows affected by selection can be detected by outlier analyses for each of the approaches, but effect size needs to be high to reach genome-wide significance when millions of markers are tested. As expected for all selection tests, we did not detect such a single strong association for any specific genomic region.

As previously shown, correction for multiple testing in genome-wide association studies on a high number of SNPs is strong and it is therefore unlikely that associated SNPs with weak or intermediate effect sizes show up as strong outliers in genome-wide association studies based on limited sample sizes (Wang et al., 2011). True signals of weak associations at specific SNPs are expected to be embedded among false signals of many random SNPs. Thus, it is recommended to further analyse the top associated signals irrespective of their genome-wide significance as single sites. In particular, tests for enrichment of functionally-related gene sets can be informative if coselection of functions or pathways is suspected (Fridley & Biernacka, 2011; Mooney & Wilmot, 2015). We found a strong enrichment for genes associated with synapses for the urban-rural association analysis. The same enrichment was observed in the other independent tests, although not significant: synaptic terms were among the top 10, top one and top seven strongest gene ontology (GO) enrichment terms for the temporal, FID and selective sweep analysis, respectively. This is a high rank given that there were only 400 terms with “synapse” or “synaptic” in the 20,023 used GO terms. Synaptic terms were also among the top two strongest GO enrichment terms in the top gene sets of the combined analysis of the Bahia Blanca samples. The 60 different genes related to these synaptic functions are predominantly expressed in the brain and some of them serve direct excitatory (e.g., *GRIA4*), inhibitory (e.g., *GABBR2*, *GABRA1*), accessory (e.g., *IL1RAPL1*) or modifying (e.g. *GRIP2*) synaptic functions (NCBI, gene db). Many of these genes have been related to personality, behaviour control, memory and cognitive/learning functions (e.g., *ANKS1B*, *APP*, *CACNA1C*, *CADM2*, CADPS2, *CTNNA2*, *CHAT*, *DLG2*, *GRIA4*, *IL1RAPL1*, *PLCB4*, *SLC5A7*, *TNIK*) (NCBI, gene db). In general, synapse proteome diversity in

vertebrates plays a fundamental role in generating complexity in the behavioural repertoire and molecular manipulations of its numerous constituents show changes in various aspects of behaviour (Grant, 2016).

Strong significant enrichment was also found for genes associated with neuron projection development in the combined analysis for selection in the Bahia Blanca sample. Terms related to neuron projection were among the top four, top two, top 17 and top 41 strongest GO enrichment terms for the urban-rural habitat, temporal, FID and selective sweep analysis, respectively, although these enrichments were not significant. Again, this is a high rank given that there were only 28 terms with "neuron projection" in the 20,023 used GO terms. The 44 genes of this enrichment group are also predominantly expressed in the brain and often have been related to cognitive and motivational functions or learning/memory (e.g., *CNTNAP2*, *COL25A1*, *CTNNA2*, *DCLK1*, *NCAM1*, *NRXN3*, *SEMA5A*, *TNIK*) (NCBI, gene db).

The two main GO enrichment terms "Synapse" and "Neuron projection development" were correlated in our gene data set, because synapses are part of neuron projections: 32% (31 of 98) of the genes within these enriched functional groups annotated to both GO terms. In general, both axonic/dendritic and synapse development can modify brain connectivity and signalling and thus can influence behavioural variation. A genome-wide association meta-analysis in humans on cognitive ability and other linked traits such as personality revealed many natural polymorphisms in genes functionally enriched for synapse and neurogenesis (Lam et al., 2017). The level of natural sequence variation in cortical genes has also been linked to individual connectivity of the human brain (Xin et al., 2019). Another study reported an association between an individual polymorphism and connectivity in brain networks (Harneit et al., 2019). We thus suggest that some, if not all, of the listed genes acting in synapses and neuron projections may play a role in the modulation of brain connectivity/signalling and consequently in cognitive function and motivational behavioural control important for early adaptation to the new urban environment. In particular, the six enriched genes (*SH3KBP1*, *TNIK*, *PRKN*, *CTNNA2*, *GPM6A*, *SDK1*) shared between the urban-rural and combined Bahia Blanca association analyses might warrant further investigation (Figure 2, highlighted in Figure 1).

We considered *a priori* the two candidate genes *DRD4* and *SERT*, both of which are known to show allele frequency changes during invasions or urban colonizations (van Dongen et al., 2015; Mueller et al., 2013, 2014, 2017). The genomic regions of both genes did not harbour genome-wide significant signals, but the window across the promoter and potential 5′UTR region of the *SERT* (also *SLC6A4*) gene was nominal significant in the urban-rural habitat association analysis. This window is close (distance of 3,425 bp) to the homologous region of the *SERT* microsatellite analysed in blackbirds (Mueller et al., 2013). In urban blackbird populations the major allele frequencies at this microsatellite were in general lower than in rural ones (Mueller et al., 2013). The homologue of the well-known human indel 5-HTTLPR, however, lies within the significant window (Lesch et al., 1996). This promoter polymorphism has been linked to variation in anxiety, harm avoidance, novelty seeking and stress sensitivity (Canli & Lesch, 2007), aggression (Craig & Halton 2009), dominance (Miller-Butterworth, Kaplan, Barmada, Manuck, & Ferrell, 2007) and vigilance and cognitive functions (Canli & Lesch, 2007; Homberg & Lesch, 2011). The diversity of behavioural traits affected may partly be explained by common central brain activity patterns. Indeed, up to 10% of the variance in amygdala activation, which plays a role in processing memory and emotional reactions underlying voluntary behaviour, was explained by the human *SERT* promoter polymorphism (Munafo, Brown, & Hariri, 2008). The test parameters of *SERT* windows scored relatively high for all other selection tests (standardised test parameters >0.8), but were not significant potentially due to low power.

In summary, gene-set analyses of the burrowing owl genome and a candidate gene analysis revealed evidence for adaptation during the process of urbanization. Our data support the idea that specific genes acting in synapses and neuron projections with cognitive and emotional behavioural control functions are important targets of early selection in the new urban environment.

ACKNOWLEDGEMENTS

We thank A. Luna, A. Palma and N. Rebolo for field work and three anonymous reviewers for constructive comments. This work was supported by the Max Planck Society (to B.K.), MINECO, Spain (projects CGL2012-31888 and CGL2015-71378-P), Fundación Repsol, and a Severo Ochoa microproyecto award from Estación Biológica de Doñana (the latter three to M.C. and J.L.T.).

AUTHOR CONTRIBUTIONS

M.C., and J.L.T. initiated and maintained the long-term study of burrowing owls in Argentina. M.C., J.L.T., J.C.M., and B.K. contributed to study conception and design. H.K., and S.B. assembled the owl genome and collected the genomic variant data. J.C.M. performed all data analyses. J.C.M., M.C., J.L.T., and B.K. drafted the manuscript. All authors provided comments and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The burrowing owl genome browser can be found at http://public-genomes-ngs.molgen.mpg.de. The NCBI *Athene cunicularia* Annotation Release 100 from the NCBI Eukaryotic Genome Annotation Pipeline can be found in the NCBI genome db. The variant call format (VCF) file with all filtered SNPs and genotypes are deposited at the NCBI archive under bioproject Number PRJNA431202 and the individual population affiliations and FID phenotypes in the Data S1.

ORCID

Jakob C. Mueller https://orcid.org/0000-0001-6676-7595

REFERENCES

Abney, M. (2015). Permutation testing in the presence of polygenic variation. *Genetic Epidemiology*, *39*, 249–258. https://doi.org/10.1002/gepi.21893

Adkins-Regan, E. (1996). Neural and hormonal mechanisms of behavior: Physiological causes and consequences. In L. Houck, & L. Drickamer (Eds.), *Foundations of animal behavior* (pp. 389–405). Chicago, IL: University of Chicago Press.

Alberti, M., Correa, C., Marzluff, J. M., Hendry, A. P., Palkovacs, E. P., Gotanda, K. M., … Zhou, Y. (2016). Global urban signatures of phenotypic change in animal and plant populations. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 8951–8956.

Canli, T., & Lesch, K. P. (2007). Long story short: The serotonin transporter in emotion regulation and social cognition. *Nature Neuroscience*, *10*, 1103–1109. https://doi.org/10.1038/nn1964

Carrete, M., Martínez-Padilla, J., Rodríguez-Martínez, S., Rebolo-Ifrán, N., Palma, A., & Tella, J. L. (2016). Heritability of fear of humans in urban and rural populations of a bird species. *Scientific Reports*, *6*, 31060. https://doi.org/10.1038/srep31060

Carrete, M., & Tella, J. L. (2011). Inter-individual variability in fear of humans and relative brain size of the species are related to contemporary urban invasion in birds. *PLoS ONE*, *6*, e18859. https://doi.org/10.1371/journal.pone.0018859

Carrete, M., & Tella, J. L. (2013). High individual consistency in fear of humans throughout the adult lifespan of rural and urban burrowing owls. *Scientific Reports*, *3*, 3524. https://doi.org/10.1038/srep03524

Carrete, M., & Tella, J. L. (2017). Behavioral correlations associated with fear of humans differ between rural and urban burrowing owls. *Frontiers in Ecology and Evolution*, *5*, 54. https://doi.org/10.3389/fevo.2017.00054

Chamberlain, D. E., Cannon, A. R., Toms, M. P., Leech, D. I., Hatchwell, B. J., & Gaston, K. J. (2009). Avian productivity in urban landscapes: A review and meta-analysis. *Ibis*, *151*, 1–18. https://doi.org/10.1111/j.1474-919X.2008.00899.x

Craig, I. W., & Halton, K. E. (2009). Genetics of human aggressive behavior. *Human Genetics*, *126*, 101–113.

deLeeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized gene-set analysis of GWAS data. *PLOS Computational Biology*, *11*, e1004219. https://doi.org/10.1371/journal.pcbi.1004219

DeCandia, A. L., Brzeski, K. E., Heppenheimer, E., Caro, C. V., Camenisch, G., Wandeler, P., … vonHoldt, B. M. (2019). Urban colonization through multiple genetic lenses: The city-fox phenomenon revisited. *Ecology and Evolution*, *9*, 2046–2060. https://doi.org/10.1002/ece3.4898

Donihue, C. M., & Lambert, M. R. (2015). Adaptive evolution in urban ecosystems. *Ambio*, *44*, 194–203. https://doi.org/10.1007/s13280-014-0547-2

Elphick, M. R., Mirabeau, O., & Larhammar, D. (2018). Evolution of neuropeptide signalling systems. *Journal of Experimental Biology*, *221*, jeb151092. https://doi.org/10.1242/jeb.151092

Evans, K. L., Hatchwell, B. J., Parnell, M., & Gaston, K. J. (2010). A conceptual framework for the colonization of urban areas: The blackbird *Turdus merula* as a case study. *Biological Reviews*, *85*, 643–667.

Evans, K. L., Newton, J., Gaston, K. J., Sharp, S. P., McGowan, A., & Hatchwell, B. J. (2012). Colonisation of urban environments is associated with reduced migratory behaviour, facilitating divergence from ancestral populations. *Oikos*, *121*, 634–640. https://doi.org/10.1111/j.1600-0706.2011.19722.x

Fridley, B. L., & Biernacka, J. M. (2011). Gene set analysis of SNP data: Benefits, challenges, and future directions. *European Journal of Human Genetics*, *19*, 837–843. https://doi.org/10.1038/ejhg.2011.57

Garroway, C. J., & Sheldon, B. C. (2013). Urban behavioural adaptation. *Molecular Ecology*, *22*, 3430–3432. https://doi.org/10.1111/mec.12351

Gogarten, S. M., Sofer, T., Yu, C., Brody, J. A., Thornton, T. A., … Conomos, M. P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics*, *35*, 5346–5348. https://doi.org/10.1093/bioinformatics/btz567

Grant, S. G. N. (2016). The molecular evolution of the vertebrate behavioural repertoire. *Philosophical Transactions of the Royal Society B*, *371*, 20150051. https://doi.org/10.1098/rstb.2015.0051

Grote, S. (2018). GOFUNCR: Gene ontology enrichment using FUNC. R package version 1.2.0. https://doi.org/10.18129/b9.bioc.gofuncr

Harneit, A., Braun, U., Geiger, L. S., Zang, Z., Hakobjan, M., Donkelaar, M. M. J., … Tost, H. (2019). MAOA-VNTR genotype affects structural and functional connectivity in distributed brain networks. *Human Brain Mapping*, *40*, 5202–5212. https://doi.org/10.1002/hbm.24766

Harris, S. E., & Munshi-South, J. (2017). Signatures of positive selection and local adaptation to urbanization in white-footed mice (*Peromyscus leucopus*). *Molecular Ecology*, *26*, 6336–6350.

Homberg, J. R., & Lesch, K. P. (2011). Looking on the bright side of serotonin transporter gene variation. *Biological Psychiatry*, *69*, 513–519. https://doi.org/10.1016/j.biopsych.2010.09.024

Johnson, M. T. J., & Munshi-South, J. (2017). Evolution of life in urban environments. *Science*, *358*, eaam8327. https://doi.org/10.1126/science.aam8327

Kempenaers, B., Borgstrom, P., Loes, P., Schlicht, E., & Valcu, M. (2010). Artificial night lighting affects dawn song, extra-pair siring success, and lay date in songbirds. *Current Biology*, *20*, 1735–1739. https://doi.org/10.1016/j.cub.2010.08.028

Kern, A. D., & Schrider, D. R. (2016). discoal: Flexible coalescent simulations with selection. *Bioinformatics*, *32*, 3839–3841. https://doi.org/10.1093/bioinformatics/btw556

Kern, A. D., & Schrider, D. R. (2018). diploS/HIC: An updated approach to classifying selective sweeps. *G3: Genes, Genomes, Genetics*, *8*, 1959–1970.

Lam, M., Trampush, J. W., Yu, J., Knowles, E., Davies, G., Liewald, D. C., … Lencz, T. (2017). Large-scale cognitive GWAS meta-analysis reveals tissue-specific neural expression and potential nootropic drug targets. *Cell Reports*, *21*, 2597–2613. https://doi.org/10.1016/j.celrep.2017.11.028

Langmead, B., & Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359. https://doi.org/10.1038/nmeth.1923

Lesch, K.-P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., … Murphy, D. L. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, *274*, 1527–1531. https://doi.org/10.1126/science.274.5292.1527

Martinez, G., Baladron, A. V., Cavalli, M., Bo, M. S., & Isacch, J. P. (2017). Microscale nest-site selection by the burrowing owl (*Athene cunicularia*) in the pampas of Argentina. *The Wilson Journal of Ornithology*, *129*, 62–70.

Marzluff, J. M. (2016). A decadal review of urban ornithology and a prospectus for the future. *Ibis*, *159*, 1–13. https://doi.org/10.1111/ibi.12430

Miller-Butterworth, C. M., Kaplan, J. R., Barmada, M. M., Manuck, S. B., & Ferrell, R. E. (2007). The serotonin transporter: Sequence variation in *Macaca fascicularis* and its relationship to dominance. *Behavior Genetics*, *37*, 678–696. https://doi.org/10.1007/s10519-007-9162-3

Mooney, M. A., Nigg, J. T., McWeeney, S. K., & Wilmot, B. (2014). Functional and genomic context in pathway analysis of GWAS data. *Trends in Genetics*, *30*, 390–400. https://doi.org/10.1016/j.tig.2014.07.004

Mooney, M. A., & Wilmot, B. (2015). Gene set analysis: A step-by-step guide. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *168*, 517–527. https://doi.org/10.1002/ajmg.b.32328

Mueller, J. C., Edelaar, P., Baños-Villalba, A., Carrete, M., Potti, J., Blas, J., … Kempenaers, B. (2017). Selection on a behaviour-related gene during the first stages of the biological invasion pathway. *Molecular Ecology*, *26*, 6110–6121. https://doi.org/10.1111/mec.14353

Mueller, J. C., Edelaar, P., Carrete, M., Serrano, D., Potti, J., Blas, J., … Tella, J. L. (2014). Behaviour-related DRD4 polymorphisms in invasive bird populations. *Molecular Ecology*, *23*, 2876–2885.

Mueller, J. C., Kuhl, H., Boerno, S., Tella, J. L., Carrete, M., & Kempenaers, B. (2018). Evolution of genomic variation in the burrowing owl in response to recent colonization of urban areas. *Proceedings of the Royal Society B: Biological Sciences*, *285*, 20180206.

Mueller, J. C., Partecke, J., Hatchwell, B. J., Gaston, K. J., & Evans, K. L. (2013). Candidate gene polymorphisms for behavioural adaptations during urbanization in blackbirds. *Molecular Ecology*, *22*, 3629–3637. https://doi.org/10.1111/mec.12288

Munafo, M. R., Brown, S. M., & Hariri, A. R. (2008). Serotonin transporter (5-HTTLPR) genotype and amygdala activation: A meta-analysis. *Biological Psychiatry*, *63*, 852–857. https://doi.org/10.1016/j.biopsych.2007.08.016

Munshi-South, J., Zolnik, C. P., & Harris, S. E. (2016). Population genomics of the Anthropocene: Urbanization is negatively associated with genome-wide variation in white-footed mouse populations. *Evolutionary Applications*, *9*, 546–564. https://doi.org/10.1111/eva.12357

Perrier, C., delCampo, A. L., Szulkin, M., Demeyrier, V., Gregoire, A., & Charmantier, A. (2018). Great tits and the city: Distribution of genomic diversity and gene-environment associations along an urbanization gradient. *Evolutionary Applications*, *11*, 593–613. https://doi.org/10.1111/eva.12580

Ramanan, V. K., Shen, L., Moore, J. H., & Saykin, A. J. (2012). Pathway analysis of genomic data: Concepts, methods, and prospects for future development. *Trends in Genetics*, *28*, 323–332. https://doi.org/10.1016/j.tig.2012.03.004

Ravinet, M., Elgvin, T. O., Trier, C., Aliabadian, M., Gavrilov, A., & Saetre, G. P. (2018). Signatures of human-commensalism in the house sparrow genome. *Proceedings of the Royal Society B: Biological Sciences*, *285*, 20181246. https://doi.org/10.1098/rspb.2018.1246

Rebolo-Ifrán, N., Carrete, M., Sanz-Aguilar, A., Rodríguez-Martínez, S., Cabezas, S., Marchant, T. A., … Tella, J. L. (2015). Links between fear of humans, stress and survival support a non-random distribution of birds among urban and rural habitats. *Scientific Reports*, *5*, 13723. https://doi.org/10.1038/srep13723

Rebolo-Ifrán, N., Tella, J. L., & Carrete, M. (2017). Urban conservation hotspots: Predation release allows the grassland-specialist burrowing owl to perform better in the city. *Scientific Reports*, *7*, 3527. https://doi.org/10.1038/s41598-017-03853-z

Rodríguez-Martínez, S., Carrete, M., Roques, S., Rebolo-Ifrán, N., & Tella, J. L. (2014). High urban breeding densities do not disrupt genetic monogamy in a bird species. *PLoS ONE*, *9*, e91314. https://doi.org/10.1371/journal.pone.0091314

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

United Nations, Department of Economic and Social Affairs. (2018). 2018 Revision of World Urbanization Prospects. https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html

Van der Auwera, G. A., Carneiro, M., Hartl, C., Poplin, R., Del Angel, G., & Levy-Moonshine, A. (2013). From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 11.10.1–11.10.33.

vanDongen, W. F. D., Robinson, R. W., Weston, M. A., Mulder, R. A., & Guay, P.-J. (2015). Variation at the DRD4 locus is associated with wariness and local site selection in urban black swans. *BMC Evolutionary Biology*, *15*, 253. https://doi.org/10.1186/s12862-015-0533-8

Wang, L., Jia, P., Wolfinger, R. D., Chen, X., & Zhao, Z. (2011). Gene set analysis of genome-wide association studies: Methodological issues and perspectives. *Genomics*, *98*, 1–8. https://doi.org/10.1016/j.ygeno.2011.04.006

Xin, Q., Ortiz-Teran, L., Diez, I., Perez, D. L., Ginsburg, J., El Fakhri, G., & Sepulcre, J. (2019). Sequence alterations of cortical genes linked to individual connectivity of the human brain. *Cerebral Cortex*, *29*, 3828–3835. https://doi.org/10.1093/cercor/bhy262

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yap, E.-L., & Greenberg, M. E. (2018). Activity-regulated transcription: Bridging the gap between neural activity and behavior. *Neuron*, *100*, 330–348. https://doi.org/10.1016/j.neuron.2018.10.013

Yu, G., Wang, L., Han, Y., & He, Q. (2012). CLUSTERPROFILER: An R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, *16*, 284–287. https://doi.org/10.1089/omi.2011.0118

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*, 3326–3328. https://doi.org/10.1093/bioinformatics/bts610

Zuk, M., Bastiaans, E., Langkilde, T., & Swanger, E. (2014). The role of behaviour in the establishment of novel traits. *Animal Behaviour*, *92*, 333–344. https://doi.org/10.1016/j.anbehav.2014.02.032

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: MuellerJC, Carrete M, Boerno S, Kuhl H, Tella JL, Kempenaers B. Genes acting in synapses and neuron projections are early targets of selection during urban colonization. *Mol Ecol*. 2020;00:1–10. https://doi.org/10.1111/mec.15451