Please cite this publication as follows:


Link to official URL (if available):

http://dx.doi.org/10.1007/s10592-016-0907-5

This version is made available in accordance with publishers’ policies. All material made available by CReaTE is protected by intellectual property law, including copyright law. Any use made of the contents should comply with the relevant law.

Contact: create.library@canterbury.ac.uk
Population genomics applications for conservation: the case of the tropical dry forest dweller

Peromyscus melanophrys

Rodrigo Vega¹,²,³*, Ella Vázquez-Domínguez³*, Thomas A. White¹,², David Valenzuela-Galván⁴ and Jeremy B. Searle².

¹ Section of Life Sciences, School of Human and Life Sciences, North Holmes Road, Canterbury Christ Church University, Canterbury, Kent, CT1 1QU, UK.
² Department of Ecology and Evolutionary Biology, Corson Hall, Cornell University, Ithaca, NY 14853-2701, USA.
³ Departamento de Ecología de la Biodiversidad, Instituto de Ecología, Universidad Nacional Autónoma de México (UNAM), Ciudad Universitaria, México DF 04510, México.
⁴ Centro de Investigación en Biodiversidad y Conservación (CIByC), Universidad Autónoma del Estado de Morelos (UAEM), Cuernavaca, Morelos, México.

*These authors contributed equally to this work.

Corresponding author:

Rodrigo Vega. Section of Life Sciences, School of Human and Life Sciences, North Holmes Road, Canterbury Christ Church University, Canterbury, Kent, CT1 1QU, UK.

Email: rodrigo.vega@canterbury.ac.uk (Tel: +44-1-227-782275, Fax: +44-1-227-767531)
Abstract

Recent advances in genomic sequencing have opened new horizons in the study of population genetics and evolution in non-model organisms. However, very few population genomic studies have been performed on wild mammals to understand how the landscape affects the genetic structure of populations, useful information for the conservation of biodiversity. Here, we applied a genomic approach to evaluate the relationship between habitat features and genetic patterns at spatial and temporal scales in an endangered ecosystem, the Tropical Dry Forest (TDF). We studied populations of the Plateau deer mouse *Peromyscus melanophrys* to analyse its genomic diversity and structure in a TDF protected area in the Huautla Mountain Range (HMR), Mexico based on 8,209 SNPs obtained through Genotyping-by-Sequencing. At a spatial scale, we found a significant signature of isolation-by-distance, few significant differences in genetic diversity indices among study sites, and no significant differences between habitats with different levels of human perturbation. At a temporal scale, while genetic diversity levels fluctuated significantly over time, neither seasonality nor disturbance levels had a significant effect. Also, outlier analysis revealed loci potentially under selection. Our results suggest that the population genetics of *P. melanophrys* may be little impacted by anthropogenic disturbances, or by natural spatial and temporal habitat heterogeneity in our study area. The genome-wide approach adopted here provides data of value for conservation planning, and a baseline to be used as a reference for future studies on the effects of habitat fragmentation and seasonality in the HMR and in TDF.

Keywords: Conservation genetics; Global Ecoregion; Mexico; non-model organism; rodents; SNP.
44 Introduction

45 Developments in high-throughput sequencing and computing permit analysis of thousands of DNA
46 polymorphisms and allow a vast range of population genetic and evolutionary questions to be addressed
47 in non-model organisms (Allendorf et al. 2010; Narum et al. 2013; Ellegren 2014); however, there are still
48 very few genomic studies on wild mammals (but see Miller et al. 2011; White et al. 2013). Genomic
49 studies have enormous potential to assess aspects of biodiversity conservation, providing the tools to
50 address the effects of habitat change, at fine scales, on genetic structure, diversity and local adaptation
51 (Allendorf et al. 2010). Moreover, spatial and temporal genomic studies of natural populations can help
52 understand the effects of factors like seasonality, fragmentation and human activities, among others, on
53 population genetic diversity and structure (Kettle 2014). Such information can help establish
54 conservation priorities and management plans in heterogeneous, perturbed or even well conserved
55 habitats like nature reserves. In addition, population and conservation genetics aim to understand how
56 the landscape affects the genetic structure of populations (Manel et al. 2003). The negative effects of
57 habitat loss on natural populations can result from a decrease in overall habitat availability, changing the
58 spatial organisation of resources and habitat quality within fragments (Fahrig 2003; Ezard and Travis
59 2006). Fragmentation decreases effective population size and gene flow, further eroding neutral and
60 adaptive genetic diversity of natural populations, lowering the evolutionary potential and increasing the
61 extinction risk through time (Frankham 2005).

62 The Tropical Dry Forest (TDF) is one of the most extensive tropical vegetation types in
63 Mesoamerica, of major cultural and economic importance, and with high levels of biological diversity
64 and endemism harbouring about 19% of Mesoamerican endemic fauna (Miles et al. 2006; Ceballos et al.
65 2010). The Mexican TDF has been recognised as a Global Ecoregion —the most diverse tropical dry
66 forests in the world— a habitat that is also highly endangered (WWF 2000). Habitat loss is one of the
67 most important threats for biodiversity (Fischer and Lindenmayer 2007) and the TDF has one of the
fastest rates of deforestation and land conversion of all the tropical forests, with only about 27% of the
corresponding forests remaining intact and 73% showing varying levels of alteration (Miles et al. 2006; Ceballos
et al. 2010). Habitat loss and fragmentation can also be associated with extirpation of species and
populations and declines in local species abundance, with consequent potential genetic erosion and
decrease of genetic diversity; processes that can have more subtle but no less important negative effects
on biodiversity (Frankham 2005; Dirzo et al. 2014). As a result, the maintenance of species, species
interactions and ecosystem functioning within the TDF may be significantly hindered at different spatial
and temporal scales.

It should be noted in the TDF that seasonality in addition to habitat alteration can have a
profound impact on animal survival and behaviour, and can also directly affect patterns of genetic
diversity and structure in animal species (Vázquez-Domínguez et al. 2002; Ceballos and Valenzuela 2010;
Liu et al. 2013).

Of the animals found in the TDF, rodents play a key role because they are the main seed
predators and dispersers of many plant species (Ceballos et al. 2010). Moreover, studies have also shown
that rodent species (e.g. Liomys pictus, Heteromys spp., Baiomys musculus, Peromyscus melanophrys)
show dynamic responses to habitat change, fragmentation and seasonality at the population and
community levels, with changes in abundance, biomass, density, genetic diversity, among other
characteristics (Vázquez-Domínguez et al. 1998, 1999, 2002; Vargas et al. 2012; Mussali-Galante et al.
2013; Garrido-Garduño et al. 2015).

The Plateau deer mouse Peromyscus melanophrys (Rodentia; Cricetidae) is a Mexican endemic
species, nocturnal, mostly herbivorous and with semi-arboreal habits (Álvarez-Castañeda et al. 2008); it
is abundant and widespread over a geographic range that includes a great proportion of undisturbed
habitat across 18 of the 32 Mexican states, several natural protected areas and a great variety of habitats
ranging from sea level up to 2,700 m altitude (Aragón 2005), but it is not found in highly disturbed
habitats and in urban or suburban environments. The species is short lived, likely between one and two years (Shug et al. 1991; Aragón 2005), it is easily captured, it has been considered as a biomonitor indicating environmental quality (Tovar-Sánchez et al. 2012). Individual movement is known to be, on average, less than 200 m, while home ranges vary (in other Peromyscus species) between 0.5 to 1.2 ha, which is dependant on habitat type (e.g. larges home ranges in desert environments) (Stickel 1968; MacSwiney et al. 2012). Hence, *P. melanophrys* is a suitable study system to evaluate the relationship between habitat features and genetic patterns at spatial and temporal levels (Lui et al. 2013) under a population genomics approach, both because of the good sample numbers required for these techniques (Davey et al. 2011), and because results can serve as a basis for applying this approach for management and conservation of ecologically similar but less abundant species.

Accordingly, our aim was to study the population genomics of this species, distributed both in undisturbed and disturbed habitats in one of the largest protected areas of TDF in Mexico, the Sierra de Huautla Biosphere Reserve (SHBR) (CONANP 2005), which is embedded within the Huautla Mountain Range (HMR), an extensive high elevation area dominated by TDF vegetation. Previous genetic studies show that *P. melanophrys* from this area belongs to a well-defined phylogeographic lineage distributed along the Balsas River basin (Castañeda-Rico et al. 2014). We used high-throughput sequencing to develop Single Nucleotide Polymorphisms (SNPs) to achieve our objectives: 1) to analyse the spatial and temporal genomic diversity and structure of *P. melanophrys* in a heterogeneous landscape within the HMR, 2) to evaluate if density and biomass varies spatially and temporally and if this correlates with the genomic variation of the species, and 3) to discuss the conservation implications of our findings and establish genetic metrics that may be used as a reference for future studies on the effects of habitat fragmentation in TDF and other natural systems. We would expect density, biomass and genetic diversity of *P. melanophrys* to be lower in disturbed vs undisturbed habitats, due to lower tree cover and lower
plant diversity and food availability, whereas we predict high genetic structure among sites across the HMR.

Materials and methods

Study region and vegetation regime

Sampling sites were located along the HMR and in the SHBR (Fig. 1), a 600 km² natural protected area characterized by a strong seasonality (Online Resource Fig. S1), and composed of a patchy landscape of well-preserved habitat (mostly TDF) intermixed with agricultural land, secondary vegetation and disturbed forest areas (CONANP 2005). Land-use changes and the disturbance history in the HMR has been long-term and gradual, and the main sources of disturbance at the TDF in the HMR are still extensive cattle raising, selective logging and the opening of agricultural lands (Maass et al. 2010). Each site was classified as disturbed (D) or undisturbed (U) based on vegetation regime (see below) as follows:

Axuchitlán (Axu-D, Axu-U), Quilamula (Qui-D, Qui-U) and Xantiopan (Xan-D, Xan-U) at the centre of the SHBR. Samples were also obtained between 2006 and 2011 from two sites from the northeast, El Limón-1 (Lim-1) and El Limón-2 (Lim-2), and from two sites in the southeast, El Salado (Sal-1) and Teotlaco (Teo-1) (with no information on vegetation regime).

To define the vegetation regime (i.e. disturbed and undisturbed) for the sites in the central part of the eastern HMR, satellite images for 1 km² sampling sites were used and the proportional tree coverage with different vegetation type within each 1 km² grid cell was assessed. Disturbed sites were pre-identified as having a low percentage (60% or less) of TDF coverage (and less than 40% of it considered to be well preserved TDF) and more than 30% of tree coverage represented by other vegetation types including secondary vegetation, agricultural fields and grasslands. Disturbed sites also have a high foraging activity of livestock. Undisturbed sites were pre-defined as having 80% or more of a predominantly dense, continuous and heterogeneous TDF coverage (with at least 40% of it considered to
be well preserved TDF), and a low percentage (<10%) of ground coverage represented by other vegetation types including agricultural lands. Undisturbed sites have a low foraging activity of livestock. Pairs of nearby disturbed and undisturbed sites that contrasted the most in terms of vegetation coverage were selected for rodent trapping (distance between sites in a pair averaged 2 km).

To evaluate if the pre-defined disturbed and undisturbed sites differed significantly in terms of biological diversity of plants, all woody plants with a diameter at breast height above 1 cm in 10 Gentry type transects of 50x2 m (Gentry 1982) were taxonomically identified and Shannon's diversity index (H') was obtained for each site. The availability of fleshy fruits was also estimated for pre-determined disturbed and undisturbed sites based on the taxonomical identification of plants on sampling transects (De León-Ibarra 2005).

Trapping and sampling

Rodent trapping was performed using Sherman traps baited with oatmeal and peanut butter, in 100 x 100 m trapping grids (100 traps per grid, spaced every 10 m) at 10 sites across the eastern part of the HMR from 2002 to 2011; rodent density values (individuals/ha) were obtained from sampling sessions for all sites based on the Minimum Number Known Alive (MNKA; Krebs, 1966), while rodent biomass (g/ha) was derived from the same procedure. A total of 153 individuals of *P. melanophrys* were captured, standard body measures were taken, and a tissue sample (toe clip) was obtained following ethical guidelines (Sikes and Gannon 2011) (see Tables 1 and 2 for sample sizes, and Appendix S1 for density and biomass values; values presented are averaged across sampling sessions for each period or site).

Genotyping-by-Sequencing (GBS) and SNP genotyping

DNA extraction was performed using the DNeasy Blood and Tissue DNA extraction kit (Qiagen) and all samples were electrophoresed in agarose gels to assess DNA quality. Double stranded DNA was
quantified with a Qubit 2.0 fluorometer (Qiagen). Extracted DNA was sent to the Cornell Institute for Genomic Diversity to conduct Genotyping-by-Sequencing (GBS) (Elshire et al. 2011). GBS is a simple technique for constructing reduced representation genomic libraries for the Illumina sequencing platform. DNA from each individual was separately digested using the restriction enzyme PstI (recognition site: CTGCAG, overhang: TGCA-3'; New England Biolabs). Given the nature of the DNA samples, and to obtain sufficient coverage (5X minimum), three genomic libraries were sequenced and results per sample were pooled. SNP calling resulted in 103,286 potential SNPs. After filtering potential SNPs using the conservative approach applied in White et al. (2013), we obtained 8,209 loci which could be confidently called in at least 90% of individuals (see Appendix S1 for further details).

Genomic diversity and structure

For the spatial analysis, the total sample was grouped into: 1) ‘Sites’ according to the locality where samples were taken from (namely Axu-U, Axu-D, Qui-U, Qui-D, Xan-U, Xan-D, Lim-1, Lim-2, Sal-1 and Teo-1), and 2) ‘Habitat Types’ by pooling samples from the undisturbed and disturbed sites across the HMR, respectively. Sample sizes are shown in Online Resource Table S1. For the temporal analysis (2002-2006), the sample from the centre of the Sierra de Huautla Biosphere Reserve was grouped into: 1) ‘One-Year Cycles by Habitat Type’ covering a full breeding cycle of *P. melanophrys* by clustering a wet season (from July to October) and a dry season (from November to June the following year) resulting in four cycles for the disturbed and the undisturbed habitat types, respectively (namely T1-U, T1-D, T2-U, T2-D, T3-U, T3-D, T4-U and T4-D), and 2) ‘Wet and Dry Seasons’ analysing wet and dry seasons separately (wet seasons from July to October and dry seasons from November to June the following year, regardless of habitat type; analysed only for 2002-2006 to provide adequate sample size per cycle and season). Samples sizes are shown in Online Resource Table S2.
Standard intra-population level diversity indices were calculated, including Nei’s (1987) gene diversity that takes into account sample size, and the population mutation rate parameter theta, using Arlequin v.3.5 (Excoffier et al. 2005). Measures of allelic and private allelic richness (ranging between 1 and 2 for bi-allelic SNPs) were calculated using HP-RARE v.1.0 (Kalinowski 2005), which uses rarefaction to correct for sampling error to produce unbiased estimates. Locus-specific diversity indices were divided into the groups mentioned above for the spatial and temporal analyses, and statistical comparisons were made with JMP v.10.0 (SAS Institute Inc.).

The genetic structure within the HMR was tested by an analysis of molecular variance (AMOVA). Average inbreeding coefficients $F_{IS}$ (Slatkin 1991) and pairwise Slatkin’s linearized $F_{ST}$ values (Slatkin 1995) between sites were estimated; significance was tested with 20,000 permutations of gene copies between individuals within populations and 20,000 permutations of individual genotypes among sampling sites to construct a null distribution, respectively. We tested for a pattern of isolation-by-distance across the HMR with a Mantel test by estimating the significance of the correlation between pairwise (Ln-transformed) geographic distances and Slatkin’s linearized $F_{ST}$ values between sites. All analyses were performed with Arlequin.

The minimum number of populations that best explained the data was evaluated using Structure v.2.3.4 (Pritchard et al. 2000) and Discriminant Analysis of Principal Components (DAPC) (Jombart et al. 2010), two programs that use different methodologies to infer population structure. Structure uses a Bayesian model-based clustering method for inferring population structure using genotype data and assigning individuals to populations. Based on preliminary runs showing convergence, we selected the following parameters: admixture model (allowing for mixed ancestry), usepopinfo=0, $K=1$–10 clusters (with 10 replicates per $K$ to check for convergence), locprior=0, burn-in=150,000, numreps=300,000, inferalpha=1, inferlambda=0. The number of $K$ was inferred using Evanno’s method (Evanno et al. 2005).

Results were collated using Structure-Harvester (Earl and vonHoldt 2012); replicate samples were
aligned using CLUMPP v.1.1.2 (Jakobsson and Rosenberg 2007) and graphically displayed using Distruct v.1.1 (Rosenberg 2004). DAPC is a multivariate method for identifying genetic clusters using K-means of principal components when group priors are unknown. The K-means were run sequentially (k=1-10) to find the best supported number of clusters based on the lowest Bayesian Information Criterion (BIC). DAPC runs were performed in R (R Development Core Team) using the package adegenet (Jombart 2008).

Outlier SNPs and annotation

We performed an outlier analysis to detect SNPs putatively under selection with the programs Lositan workbench (Antao et al. 2008) and BayeScan v.2.1 (Foll and Gaggiotti 2008). We used both programs for a more robust identification because they differ in the detection method used. Outlier loci, potentially under selection, were then excluded from the analysis of population genomic diversity and structure, which assumes neutrality of the molecular markers.

Lositan is an FST-detection method based on the FDIST program (Beaumont and Nichols 1996), which evaluates the relationship between Wright’s inbreeding coefficient FST and heterozygosity under an island model of migration. The distribution created is used to identify outlier loci that have excessively high or low FST values compared to neutral expectations (Antao et al. 2008). Two runs were done with Lositan with the parameters: number of simulations=100, number of populations=10, confidence interval=0.99, False Discovery Rate (FDR)=0.05, dataset mean FST=0.0253 (FST=0.0121 for the temporal analysis). The first run using all loci resulted in several candidate outliers, which were then excluded and a new mean FST was obtained, FST=0.025 (FST=0.0104 for the temporal analysis). The second run using all loci was conducted using the new mean FST, lowering the bias on the estimation of the mean neutral FST by having removed the most extreme loci from the estimation.
BayeScan identifies outlier loci using differences in allele frequencies between populations based on the multinomial-Dirichlet model, and using a Bayesian approach to incorporate uncertainty on allele frequencies due to small sample sizes with low risk of bias. For BayeScan we used the parameters: burn-in=50,000, thinning interval=20, sample size=10,000, number of pilot runs=50, length of pilot runs=5000, total number of iterations=250,000 and number of populations=10.

The DNA sequences containing the outlier SNPs were compared with the GenBank database using BlastN v.2.2.28 (Altschul et al. 1997) to annotate them. Parameters were as follows: word_size=11; gapopen=5, gapextend=2; reward=2, penalty=-3. Loci were identified as putatively genic if they had an expect value $e \leq 1 \times 10^{-5}$ in matches to the nucleotide database. Only the first significant hit with the lowest $e$ was recorded. The significant matches were then incorporated into the UniProt Knowledgebase (UniProtKB) and only loci with identified gene ontologies and/or protein names were recorded.

Results

Vegetation regime

Woody plant $H'$ differed significantly between vegetation regimes ($H'_{undisturbed}=0.913$, $H'_{disturbed}=0.805$; $F_{1,1}=10.44$, $P<0.05$). The availability of fleshy fruits differed significantly between sites, being higher at undisturbed sites ($F_{1,1}=20.7736$, $P<0.05$).

SNP genotyping

From a total of 153 samples of $P$. melanophrys, 135 individuals were genotyped successfully in 8,209 loci with less than 10% missing data. The rest did not amplify well and were discarded from analysis. Excluding the outlier loci (see below), we analysed a total of 8,035 neutral loci. SNP data are available in NCBI Sequence Read Archive (BioProject: PRJNA297572; BioSamples: SAMN04127679- SAMN04127813).
Spatial analysis

The density and biomass of *P. melanophrys* sometimes differed between paired disturbed and undisturbed sites but did not significantly differ between the two categories overall (Fig. 2a,b; Online Resource Table S1). No significant correlations between density and biomass with any genetic diversity index were observed.

We found few significant differences in terms of genetic diversity indices among sites in the HMR (Fig. 2, Online Resource Table S1). For the pooled samples the only significant difference was for theta, which showed lower values in disturbed habitats; if the pair of sites Aux-U and Aux-D were removed from the analysis, the result was no longer significant, likely associated with the small sample size of Aux-U (*N*=8). There were significant *F*~*ST~ values in undisturbed and disturbed habitats (Online Resource Table S1). By site, there were no significant differences in heterozygosity, allelic richness and proportion of private alleles between all the central sites, although values were significantly higher than those from the eastern and northern sites (Fig. 2c,e,f). Theta values from the central sites were similar regardless of the vegetation regime, with the exception of Axu-U that had a higher value (Fig. 2d).

Pairwise genetic differences between sites were low on average (*F*~*ST~*=0.0227), with the highest values between the central with northern and eastern sites (Online Resource Table S3), while genetic differentiation (*F*~*ST~) among disturbed sites was similar to that than among undisturbed sites (0.0156 and 0.0151, respectively). Regarding AMOVA results, 93.93% of the total variation was distributed within individuals, whereas only 1.74% was between sites and 4.33% between individuals within sites. When analysed based on habitat types, AMOVA results were: 0.34% of the total variation between undisturbed and disturbed habitat types, 5.6% between individuals within habitat types and 94.1% within individuals.

Genomic differentiation showed a significant pattern of isolation-by-distance within the HMR (*F*~*ST~*=0.0227, *R*^2^=0.471, *P*<0.05) (Fig. 3). Structure analysis inferred the highest probability at Delta K=2, indicating that Quilamula and El Limón were the most differentiated populations (Fig. 4). Similarly, with
DAPC $K=1$ and $K=2$ had the lowest BIC values, indicating no population structure based on the SNP dataset and Quilamula was the only differentiated population based on group memberships.

Temporal analysis

*P. melanophrys* density and biomass varied through time in the three-year period analysed ('One-year Cycles by Habitat Type'), but these changes were not significantly different in disturbed and undisturbed habitats. In undisturbed sites density and biomass increased from T1 to T2 and decreased in T3, while in disturbed sites density and biomass decreased during the three one-year cycles (Fig. 5a,b; Online Resource Table S2). No significant correlations between density or biomass and any genetic diversity index were found.

Only marginally significant differences were observed in allelic richness through time (Fig. 5e, Online Resource Table S2); however, there was a significant increase in heterozygosity and allelic richness in disturbed habitats at T3 compared with undisturbed habitats (Fig. 5c,e). Also, no differences in theta values were observed except for an increase in T3-U and T3-D compared with the two previous cycles, a higher value at T4-U and again lower values in T4-D (Fig. 5d). The proportion of private alleles was always higher in disturbed than in undisturbed habitats in each cycle; this proportion decreased in T1 to T2 and remained low during T3 and T4 for undisturbed sites, whereas in disturbed sites the proportion oscillated during each cycle (Fig. 5e). Finally, there was a significant difference in the proportion of private alleles in T3 between undisturbed and disturbed sites (Fig. 5f).

The AMOVA results showed the same trend, with 94% of genomic variation within individuals. We found significant high $F_{IS}$ values at T1-U, T1-D and T2-U, which was unexpected since T1-U and T1-D have relatively high heterozygosity and allelic richness values (Online Resource Table S2).

When analysed by ‘Wet and Dry Seasons’ density and biomass fluctuated through time from lower values in wet seasons to higher values in dry seasons, but the differences were not significant (Fig.
Again, no significant correlations between rodent density or biomass and any genetic diversity index were observed. However, significant changes in genetic diversity indices were observed between seasons (Fig. 6c,e,f, Online Resource Table S2), in which values oscillated (increased/decreased) during the different wet and dry seasons. AMOVA results showed 94.1% of genomic variation within individuals, while the dry seasons 2002-2003 and 2003-2004 showed significantly high $F_{IS}$ values (Online Resource Table S2).

Outlier SNPs and annotation

Out of the 8,209 loci, a total of 174 outliers were detected based on unusual $F_{ST}$ values, which were considered as ‘putatively under selection’. With Lositan, 104 outlier loci had lower than expected $F_{ST}$ values and were putatively under balancing selection, and 53 outlier loci had higher than expected $F_{ST}$ values, putatively under directional selection. With BayeScan, 28 outlier loci were detected, all of which had higher than expected $F_{ST}$ values, 11 of which were also detected by Lositan.

From the 174 outliers, 79 sequences had an $e \leq 1 \times 10^{-5}$ and several loci had significant hits in GenBank. The gene functions of the identified sequences were diverse, some unknown or uncharacterised (Online Resource Table S4). Under balancing selection, the most notable loci were related with intracellular signal transduction, cell chemotaxis, notch signalling pathway (cell signalling), kinase activity, B-cell proliferation, cell differentiation and ATP binding activity. Under directional (positive selection), these were related with post-translational modification of proteins (geranylgeranylation), DNA, RNA and protein binding, humoral immune response and calcium ion transport.

Discussion
We studied the spatial and temporal effects of habitat perturbation and population fluctuations, at fine scales, on the genetic structure and diversity of a small mammal within a Tropical Dry Forest (TDF), using a population genomic approach to generate novel SNP loci. Our study is also relevant in terms of biodiversity conservation: we analysed the neutral genomic diversity and structure in habitats with two different levels of perturbation, characterised outlier loci potentially under selection, and we focused on a non-model organism, *Peromyscus melanophrys*, an important and abundant species of rodent within this endangered vegetation type, closely tied to vegetation dynamics through seed removal and dispersal (Ceballos et al. 2010). Focusing conservation efforts not only on rare species, but also on those more common ones, does ensure the retention of key ecological and functional roles in ecosystems (Gaston 2010; Lindenmayer et al. 2011). Our findings provide insights into the demographic and evolutionary processes in a natural rodent population, but they also have broader implications for the conservation and management of biological diversity. The genomics approach and metrics we described here can serve as a basis for the study of other non-model organisms and for evaluating the effects of habitat fragmentation in natural ecosystems. The depletion of common species and drastic declines in abundance, beyond the expected seasonal dynamics, could be the first evidence of negative anthropogenic effects in an ecosystem. Moreover, common and/or abundant species like *P. melanophrys* are easier to monitor than rare and threatened species, an important factor to consider when economic resources for the monitoring and management of biodiversity in nature reserves are extremely limited.

Genetic diversity is the most basic component of biological diversity, which determines the potential of populations to adapt to changing environments and the vulnerability of species to extinction (Frankham 2005). However, the genetic diversity of Mexican rodents is mostly unknown, although a few studies have found it can show high levels (Vega et al. 2007; Castañeda-Rico et al. 2011; Vargas et al. 2012; Vázquez-Domínguez et al. 2013; Espindola et al. 2014). To the best of our knowledge, our work represents the first population genomic study using high-throughput sequencing in a Mexican endemic
rodent, and highlights the importance of developing genomic markers for non-model organisms in
conservation studies (Kettle 2014; Shafer et al. 2015).

Spatial and temporal population genomics

We found little spatial population genetic structure for *P. melanophrys* based on SNP data within the
Huautla Mountain Range (HMR), somehow surprising given the different vegetation regimes, the hilly
landscape, and the marked seasonality that have significant effects on population sizes of other rodents
in TDF. Our results contrast with other studies on vertebrates in fragmented and environmentally
heterogeneous habitats that show high genetic differentiation, low gene flow and bottleneck signatures
across the landscape, *e.g.* Habromys simulatus (Castañeda-Rico et al. 2011), Akodon azarae (Cavia et al.
2005) and Reithrodontomys spectabilis (Espindola et al. 2014). There is an increasing number of studies
on mammals showing low genetic structure and maintenance of high genetic diversity in fragmented
habitats, and no allelic diversity loss after bottlenecks, including for example the golden-crowned sifakas
(*Propithecus tattersalli*) (Quéméré et al. 2010), water voles (*Arvicola terrestris*) (Aars et al. 2006), and the
southern pygmy mouse (*Baiomys musculus*), present in the HMR, which also shows high genetic diversity
levels in undisturbed and disturbed sites (Vargas et al. 2012). Confounding factors and synergistic
interactions affect the detection of the impact of habitat fragmentation, and species with differing life
history strategies are differentially affected by habitat fragmentation (Ewers and Didham 2006).

Population genetic models predict that a few immigrants each generation could maintain levels
of genetic variability (Wright 1943). The comparable levels of genetic diversity and the small values of
genetic differentiation of *P. melanophrys* among the central areas of the HMR indicate that those
populations are not effectively isolated, suggesting that whether habitats are disturbed or undisturbed
had so far little impact on genetic diversity. The small genetic differentiation but significant isolation-by-
distance among populations and the differentiation into two clusters of *P. melanophrys*, may reflect
relatively uniform selective pressures that only vary at a more regional scale in the HMR. Although we did not test for female or male biased dispersal, the isolation-by-distance pattern may be related to individual behaviour, as shown for other species of Peromyscus, where females have on average a smaller home range than males, adult females are more territorial than adult males and juveniles disperse more readily than either adult males or adult females (Stickel 1968; Vázquez-Domínguez et al. 1999; MacSwiney et al. 2012).

While the effects of the landscape or habitat heterogeneity on genetic diversity have been studied extensively, there are still few temporal studies of genetic variation on wild rodents, and none involving SNPs as far as we are aware. This temporal information is essential in seasonal habitats, as different studies have shown: Gaines et al. (1997) found that, contrary to predictions, there were no significant genetic differences among populations of prairie voles at either high or low density phases, attributed to the fact that probably enough animals survived in the populations during the low-density phases, preserving most of the genetic variation and thus avoiding genetic bottlenecks. However, some heteromyid rodents, which have physiological adaptations associated with seasonality, do show significant genetic diversity changes in concert with population density fluctuations (e.g. Liomys pictus; Vázquez-Domínguez et al. 1999, 2002). It appears that genetic diversity of P. melanophrys in the HMR is maintained at similar levels through time, likely because population size does not change significantly with seasons as to cause seasonal population bottlenecks. Furthermore, the genetic structure observed in the HMR may remain low because individuals can move unrestricted from different locations and quickly repopulate the TDF, even after a drastic population bottleneck (e.g. Cavia et al. 2004). Behavioural and physiological adaptations to seasonal fluctuations and habitat disturbances in P. melanophrys may also allow it to survive in TDFs, without causing significant changes in genetic diversity through time or population structure (Vázquez-Domínguez et al. 1998, 1999).
Signals of loci under selection

Anthropogenic activities and marked seasonality, such as those present in the HMR, can generate selection pressures and phenotypic plastic responses on natural populations (Palumbi 2001; Hendry et al. 2008). For example, this could be through higher predation levels and increased exposure to UV light due to reduced vegetation cover, release of agrochemical contaminants that have adverse effects across different trophic levels, changes in food and lower water availability, among others.

Based on our results, we did not find evidence to suggest that the type of habitat disturbance and seasonal environmental fluctuations in the HMR have affected the population genomic structure of *P. melanophrys*. This suggests that *P. melanophrys* is an environmentally tolerant species, in contrast with other generalist species inhabiting TDF that show genetic changes directly associated with seasonality and habitat heterogeneity, *e.g.* *L. pictus* (Vázquez-Domínguez et al. 1998, 1999, 2002; Garrido-Garduño et al. 2015) and *B. musculus* (Vargas et al. 2012). *F*\textsubscript{ST} outlier methods have a high rate of false positives (Bierne et al. 2013; Fourcade et al. 2013), but here we used a conservative approach and consider all outliers as candidate loci under selection, requiring subsequent detailed investigation. A gene expression study based on the genes detected under putative selection would be especially informative about the physiological adaptations to environmental stress tolerance. Other factors could also affect the genetic traits of this species, for instance, heavy metal contamination from mining activities, a very localized (for now) human activity in the HMR, for which there is already evidence of impact (Mussali-Galante et al. 2013). Using a similar genomic approach may uncover other putative loci under selection in *P. melanophrys* due to heavy metal contamination.

Conservation implications

Our results provide insights into the demographic and evolutionary processes in a natural population of rodent, showing low genetic differences through space and time despite previous ecological studies.
indicating potential differences between disturbed and undisturbed sites in the HMR (Cadena, 2003; D. Valenzuela-Galván, unpublished data). In accordance, we can propose that *P. melanophrys* in the HMR be considered and managed as a single population, with high levels of genetic diversity, low genetic differentiation and with a significant isolation-by-distance pattern evident among the most distant sites. Our results are encouraging in the sense that anthropological activities in the HMR have not yet impacted negatively the genetic diversity and structure of *P. melanophrys*, supporting the notion that the conservation efforts and the management activities of the SHBR are adequate at least for this species.

For the HMR in particular, we would expect to find similar genetic patterns in other relatively generalist species of small mammals inhabiting the reserve, like *L. irroratus, Oligoryzomys fulvescens, Oryzomys couesi, P. levipes, Perognathus flavus, R. megalotis, Sciurus aureogaster* and *Sigmodon hispidus*. Indeed, comparable results have been found in *B. musculus* using other genetic markers (Vargas et al. 2012). Mammals with more specialised habitats and/or diets would be expected to suffer more from habitat loss and fragmentation, for example the Mexican endemic marsupial *Tlacuatzin canescens* that has semi-arboreal habits, *Cryptotis mexicana* with a preference for primary forest and damp, grassy areas bordering streams or orchards, *Lepus callotis* with a preference for grassy areas and sensitive to overgrazing and other human-induced disturbances, *Hodomys alleni* with a preference for lower hill slopes and rocky hillsides, and *R. fulvescens*, a small rodent with a preference for grassy fields intermixed with shrubs; however, because these species are either rare or less common than *P. melanophrys* it would be difficult to monitor population density and analyse their genetic diversity and structure effectively. The population genetic metrics we described may be used as a reference for future studies on the effects of habitat fragmentation and seasonality in TDF and other natural systems. It is important to highlight that as more genomic studies with non-model organisms are performed the
better we would be able to explain ecological and evolutionary patterns (Allendorf et al. 2010; Kettle 2014; Shafer et al 2015).

The purpose of biosphere reserves is to promote sustainable development through research and monitoring ecosystems while safeguarding biological diversity (Batisse 2003). Therefore, the research approach followed here and our results are valuable for conservation in TDFs, both in terms of genetic monitoring of a natural population (Schwartz et al. 2007) and for establishing baseline information useful for taking scientifically informed decisions about management practices in the HMR.

Acknowledgements

Computations were performed using the Cornell University Biotechnology Resource Center (BRC) and Bioportal (University of Oslo). Trapping was done under permit FAUT-0251 extended to DVG. Valentino Sorani provided data for analysis of satellite images. The Centro de Investigación en Biodiversidad y Conservación, Universidad Autónoma del Estado de Morelos (CIByC-UAEM) provided logistical facilities. DVG thanks field assistance of Martín Cadena, Edgard Mason, Evodio Rendón and Amantina Lavalle (CIByC-UAEM). We acknowledge Robert Elshire, Sharon E. Mitchell and Charlotte Acharya (Institute for Genomic Diversity, Cornell University) for their help on Genotyping-by-Sequencing.

Compliance with Ethical Standards

Funding: This project was partly funded by the Cornell Center for Comparative and Population Genomics (3CPG), Consejo Nacional de Ciencia y Tecnología (CONACyT; J3490-V), Secretaría de Medio Ambiente-CONACyT (C01-0794) and Programa Volkswagen “Por Amor al Planeta”. RV was funded by the Programa de Becas Postdoctorales, Universidad Nacional Autónoma de México (UNAM). EVD acknowledges financial support from the Instituto de Ecología, UNAM.

Conflict of Interest: The authors declare that they have no conflict of interest.
References


CONANP (2005) Programa de conservación y manejo Reserva de la Biosfera Sierra de Huautla. Comisión Nacional de Áreas Naturales Protegidas, México


De León-Ibarra MA (2005) Fenología de especies de plantas con frutos carnosos y disponibilidad espacial y temporal de este recurso en la reserve de la biosfera Sierra de Huautla: Implicaciones para los vertebrados. Dissertation, Universidad Autónoma del Estado de Morelos


Foll M, Gaggiotti OE (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977-993


Palumbi SR (2001) Humans as the world’s greatest evolutionary force. Science 293:1786-1790


Stickel LF (1968) Home range and travels. In: King JA (ed) Biology of Peromyscus (Rodentia), Special Publication No. 2. The American Society of Mammalogists, USA, pp 373-411


Electronic supplementary material The online version of this article (doi: XXXX/XXXX) contains supplementary material, which is available to authorized users.

Figure S1. Annual and monthly precipitation and temperature mean values between 1972 and 2008 in the Huautla Mountain Range. Values obtained from the meteorological station in Huautla, Mexico (Latitude: 18° 26´ 35´´, Longitude: 99° 00´ 59´´, Altitude: 968.14 m). The shaded areas correspond to the sampling period used for the temporal analysis of genetic diversity and structure (2002-2006).

Appendix S1. Extended Materials and Methods: Tropical Dry Forest (TDF) vegetation in the Huautla Mountain Range (HMR), Disturbed and undisturbed sites classification, and Genotyping-by-Sequencing (GBS): SNP genotyping and calling

Table S1. Spatial values of genomic diversity, density and biomass.

Table S2. Temporal values of genomic diversity, density and biomass.

Table S2. Pairwise genetic and geographic distances.

Table S4. BlastN and UniProtKB results for outlier loci.
Figure Legends

Fig. 1 a) Geographic distribution of *Peromyscus melanophrys* in Mexico (grey shaded area); b) location of the Huautla Mountain Range (HMR; dark shaded area) and of the Sierra de Huautla Biosphere Reserve (SHBR; state of Morelos, Mexico; grey shaded area); c) sampling localities in the eastern portion of the HMR as follows: 1 = Qui-U, 2 = Qui-D, 3 = Axu-D, 4 = Axu-U, 5 = Xan-D, 6 = Xan-U, 7 = Lim-1, 8 = Lim-2, 9 = Sal-1, 10 = Teo-1, where Axu = Axuchitlán, Qui = Quilamula, Xan = Xantiopan, Lim = El Limón, Sal = El Salado, Teo = Teotlaco; U and D mean undisturbed and disturbed habitats, respectively

Fig. 2 Population-level metrics and spatial genomic diversity values for *Peromyscus melanophrys* in the Huautla Mountain Range, Mexico. (A) density, (B) biomass, (C) expected heterozygosity (*H*<sub>E</sub>), (D) population mutation rate parameter (theta), (E) allelic richness and (F) proportion of private alleles. Density and biomass not recorded for two sites from the northeast, El Limón-1 (Lim-1) and El Limón-2 (Lim-2), and from two sites in the southeast, El Salado (Sal-1) and Teotlaco (Teo-1). Error bars represent 95% confidence intervals. Significance levels among groups indicated with letters A to E. NS = not significant, * = significant (among habitat types)

Fig. 3 Isolation-by-distance showing the relationship between pairwise (Ln-transformed) geographic distances and genetic differentiation values (Slatkin’s linearized *F*<sub>ST</sub>) between sites for *Peromyscus melanophrys* in the Huautla Mountain Range

Fig. 4 Bayesian analysis of *Peromyscus melanophrys* in the Huautla Mountain Range based on Single Nucleotide Polymorphisms (Structure plot). Each vertical line represents an individual partitioned into
one or two colour segments, indicating the individual membership for the inferred number of populations ($K = 2$). The black lines separate the sampling sites that are indicated under the plot.

**Fig. 5** Temporal population-level metrics and genomic diversity indices for ‘One-year Cycles by Habitat Type’ for *Peromyscus melanophrys* in the Huautla Mountain Range, Mexico. (A) density, (B) biomass, (C) heterozygosity ($H_e$), (D) population mutation rate parameter (theta), (E) allelic richness, and (F) proportion of private alleles. Density and biomass not recorded for T4. Black and dotted lines represent undisturbed and disturbed sites, respectively. Error bars represent 95% confidence intervals. Significance levels among groups indicated with letters A to D. NS = not significant.

**Fig. 6** Temporal population-level metrics and genomic diversity indices for ‘Wet and Dry Seasons’ for *Peromyscus melanophrys* in the Huautla Mountain Range, Mexico. (A) density, (B) biomass, (C) heterozygosity ($H_e$), (D) population mutation rate parameter (theta), (E) allelic richness, and (F) proportion of private alleles. Density and biomass not recorded for Wet-5 and Dry-05/06. Error bars are 95% confidence intervals. Significance levels among groups indicated with letters A to E. NS = not significant.
Figures

Figure 1:
Figure 2:
Figure 3:
Figure 4:
Figure 5:
Figure 6: