


# SCIENTIFIC REPORTS



OPEN

## Landscape determinants of fine-scale genetic structure of a small rodent in a heterogeneous landscape (Hluhluwe-iMfolozi Park, South Africa)

Received: 22 December 2015

Accepted: 13 June 2016

Published: 13 July 2016

Isa-Rita M. Russo<sup>1</sup>, Catherine L. Sole<sup>2</sup>, Mario Barbato<sup>1,†</sup>, Ullrich von Bramann<sup>3</sup> & Michael W. Bruford<sup>1</sup>

Small mammals provide ecosystem services, acting, for example, as pollinators and seed dispersers. In addition, they are also disease reservoirs that can be detrimental to human health and they can also act as crop pests. Knowledge of their dispersal preferences is therefore useful for population management and landscape planning. Genetic data were used alongside landscape data to examine the influence of the landscape on the demographic connectedness of the Natal multimammate mouse (*Mastomys natalensis*) and to identify landscape characteristics that influence the genetic structure of this species across a spatially and temporally varying environment. The most significant landscape features shaping gene flow were aspect, vegetation cover, topographic complexity (TC) and rivers, with western facing slopes, topographic complexity and rivers restricting gene flow. In general, thicket vegetation was correlated with increased gene flow. Identifying features of the landscape that facilitate movement/dispersal in *M. natalensis* potentially has application for other small mammals in similar ecosystems. As the primary reservoir host of the zoonotic Lassa virus, a landscape genetics approach may have applications in determining areas of high disease risk to humans. Identifying these landscape features may also be important in crop management due to damage by rodent pests.

Small mammals provide important ecosystem services<sup>1</sup>, and knowledge of their dispersal preferences is useful for population management and landscape planning<sup>2</sup>. Small mammals can regulate insect populations, act as pollinators and seed dispersers, support forest regeneration, aerate soil and provide food for carnivores and humans<sup>1</sup>. In addition, the transport of minerals from deeper layers of the soil to the surface can act as a source of nutrients for plants, and the burrowing activity of some rodent species influences surface water runoff<sup>1</sup>. Small mammals are also disease reservoirs that can be detrimental to human health and they can also act as crop pests<sup>3,4</sup>. Ecosystem services, food security and health risks posed by small mammals can only be regulated by managing habitats such that opportunities for migration are available, and this remains a key goal in population management<sup>5</sup>. It is for these reasons that it is important to understand how the landscape affects dispersal in small mammals and their associated diseases.

Dispersal in rodents is thought to be a response to a number of factors, including inbreeding avoidance, mate searching, variation in habitat quality and local competition<sup>6</sup>. However, despite its importance, information regarding dispersal patterns for many species, especially for small mammals, is often limited<sup>7</sup>.

The majority of small mammals are found in semi-isolated populations due to habitat heterogeneity<sup>8</sup>. In cases where the landscape between habitat patches severely restricts dispersal, patterns of population genetic

<sup>1</sup>Cardiff School of Biosciences, Sir Martin Evans Building, Cardiff University, Museum Avenue, Cardiff, CF10 3AX United Kingdom. <sup>2</sup>Department of Zoology and Entomology, University of Pretoria, Private Bag X20, Hatfield, 0028 South Africa. <sup>3</sup>Technische Universität Braunschweig, Institut für Geosysteme und Bioindikation, Langer Kamp 19c, Braunschweig, 38106 Germany. <sup>†</sup>Present address: Istituto di Zootecnica, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122, Piacenza, Italy. Correspondence and requests for materials should be addressed to I.M.R. (email: RussolM@Cardiff.ac.uk)

differentiation should reflect landscape permeability<sup>9</sup>. A comparative analysis of arvicoline species (lemmings, voles and muskrats) has demonstrated frequent short distance dispersal events of hundreds of meters<sup>6</sup>, however, some small mammal species are capable of dispersing over much longer distances of kilometres<sup>7</sup>.

More specifically, the Natal multimammate mouse (Rodentia: Muridae; *Mastomys natalensis*) is an opportunistic generalist, nocturnal and omnivorous. They are to some extent dependent on water but occur in areas where water is only seasonally available<sup>10</sup>. “Swimming behaviour” has also been described for this species<sup>11</sup>. The species is a major pest in sub-Saharan Africa<sup>3</sup> with a wide habitat tolerance<sup>10</sup>. These mice are almost ubiquitously distributed across the continent and when found in close contact with humans they have been linked to the transmission of disease, most significantly Lassa fever virus<sup>12,13</sup>, and with loss and contamination of food sources<sup>3</sup>. Dispersal rates and dispersal distances per generation in *M. natalensis* have been shown to be relatively high<sup>3</sup> with individuals moving over distances larger than 400 m and occupying home ranges of >1 000 square meters<sup>14</sup>.

Previous studies on small mammals<sup>15</sup> have called for additional approaches to population genetics, for example landscape genetics, to better understand dispersal in small mammal populations. Knowledge of how landscape features interact with genetic variation at both the population and individual level is required to understand gene flow and adaptation to the environment<sup>16</sup>. Landscape features can both restrict and facilitate the movement of individuals in natural populations<sup>17</sup>. Therefore, studying patterns of gene flow in relation to environmental characteristics provide indirect information on life history traits and ecological function. In addition, this provides critical information for both small mammal ecosystem services (pollinators and seed dispersers) and management<sup>18,19</sup>.

To date, small mammals have received less attention in landscape genetic studies and these studies have largely been carried out in temperate and/or developed regions, where anthropogenic effects on the landscape can be profound<sup>20,21</sup>. However, barrier effects can be just as relevant in less modified landscapes and such studies can provide relevant information for managing ecologically similar species<sup>22</sup>. Landscape modification can impact on the movement of species differently and it is therefore crucial to understand similarities among species for landscape planning<sup>23</sup>.

In this context, we investigated genetic structure and landscape heterogeneity in the Natal multimammate mouse from the Hluhluwe-iMfolozi Park (HiP), South Africa. Population structuring and division is expected to be evident in small mammals such as *M. natalensis* at small spatial scales<sup>15,24</sup>, but here we aimed to test how landscape and/or habitat features interact with dispersal. We tested the hypothesis that local-scale habitat heterogeneity (for example temperature, food availability and shelter from predators) influences patterns of genetic structure. Genetic data were therefore used alongside high-resolution landscape data to examine the influence of the landscape on the demographic connectedness of *M. natalensis* and to identify landscape characteristics that influence the genetic structure of this species across a spatially and temporally varying environment.

## Results

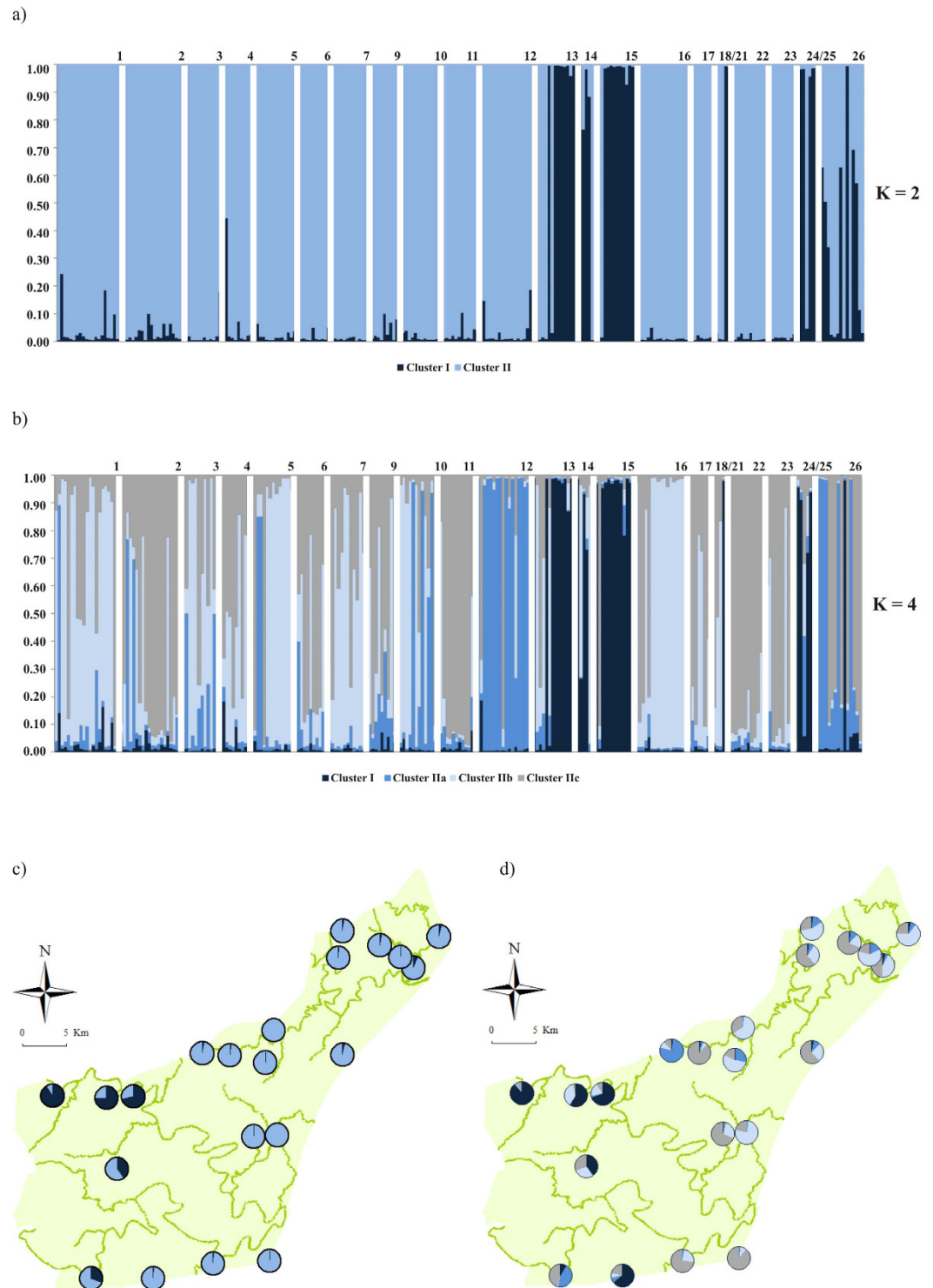
**Genetic results.** A null allele was detected for locus MH60 and we therefore excluded this locus from all further analyses. No consistent departures from Hardy-Weinberg equilibrium were detected and no consistent linkage disequilibrium was observed. The number of alleles per locus ranged from 4 to 39 and genetic clusters showed significant differentiation ( $F_{ST} = 0.025\text{--}0.124$ ). Allelic richness (254 diploid individuals) ranged from 3.99 to 38.98. Expected heterozygosity for the five loci in common with Brouat *et al.*<sup>15</sup> was also higher in the present study. The mean expected heterozygosity varied from 0.167–0.763 for these five loci.

STRUCTURE analysis indicated that the most likely value of  $\text{LnPr}(X|K)$  was  $K = 7$  however, the method of Evanno *et al.*<sup>25</sup> indicated that the most probable number of clusters of individuals was two, followed by four (see Supplementary Fig. S4). Over all runs the probability of the data increased from  $K = 1$  to  $K = 10$  with a clear tendency to asymptote (see Supplementary Fig. S4). Clustering based on the Evanno method for  $K = 2$  is shown in Fig. 1a,c. Cluster I predominantly originated from sites located south of the Black iMfolozi River and showed less evidence of admixture, whereas Cluster II showed scattered assignment across the area (Fig. 1a,c).

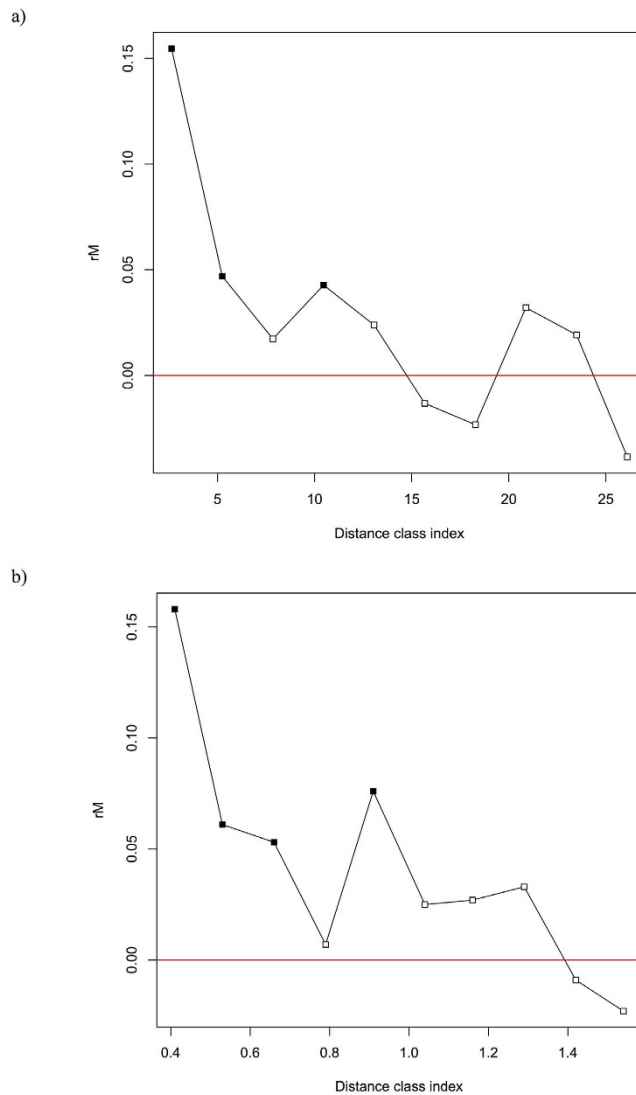
For the next highest  $\Delta K$  value (at  $K = 4$ ; see Supplementary Fig. S4) results showed no clear evidence of geographic structure among sites. Again, Cluster I predominantly originated from sites south of the Black iMfolozi River (Fig. 1b,d), whereas further sub-structuring of Cluster II into three additional clusters was evident (Fig. 1b,d). Cluster IIa occurred mostly in grid 10 and 12. Individuals with an admixture coefficient proportion equal or greater than 0.7 for Cluster IIb were more commonly found north of the White iMfolozi River ( $\chi^2 = 16.27$ ,  $P < 0.001$ ). In contrast, Cluster IIc was mostly distributed in the south of the park ( $\chi^2 = 15.43$ ,  $P < 0.001$ ). Genetic clusters were scattered across the park and therefore no strict geographic assortment was evident (Fig. 1d). These groups were also observed after the very low number of related individuals was removed from the analysis (not shown). When a full progressive partitioning method was implemented, the number of population clusters was  $K = 7$  and in agreement with the  $\text{LnPr}(X|K)$  method, although not geographically or biologically meaningful (not shown).

**Spatial autocorrelation.** After adjusting the  $P$ -value for multiple comparisons ( $P$ -value = 0.017) a Mantel correlogram showed significant and positive autocorrelation values in the first (0.026–2.63 km), second (2.63–5.24 km), third (5.24–7.85 km) and fourth (7.85–10.46 km) distance classes, exhibiting short distance structuring of genetic variability (Fig. 2a). Results showed a higher and significant genetic correlation among individuals than expected within some distance classes. When testing for association between genetic and resistance distances (resistance with a constant value of 1), significant (corrected  $P$ -value = 0.017) and positive autocorrelation values were only observed in the smaller resistance distance classes (Fig. 2b).

**Univariate and multivariate model optimisation.** Landscape resistance as a function of aspect showed the highest correlation coefficient with genetic distance, after removing the effect of the isolation-by-resistance



**Figure 1. Bayesian clustering analysis in STRUCTURE.** (a) Plot showing the individual membership coefficients for  $K=2$  when considering all rodents. Cluster I = most of the individuals south of the Black iMfolozi River, Cluster II = remainder of the individuals. (b) Plot of the individual membership coefficients for  $K=4$ . Here, Cluster II was further divided into three clusters. Cluster IIa occurred mostly in grid 10 and 12, Cluster IIb was more common north of the White iMfolozi River and Cluster IIc was mostly distributed within the southern part of the park. See maps c ( $K=2$ ) and d ( $K=4$ ) for the geographic distribution of clusters. The colours indicate different clusters and the size of the pie charts represent the frequency of occurrence for each grid sampled. Numbers (above graphs) show sampling grid numbers as indicated in Table S1. The green lines show the four major rivers in the park. Graphs were generated in Microsoft Excel for Mac v 15.19.1 (2016) and maps were generated in ArcGIS v 10.1 (<http://www.esri.com/software/arcgis/arcgis-for-desktop>).



**Figure 2.** (a) A Mantel correlogram showing a positive correlation between genetic distance (proportion of shared alleles) and Euclidean distance in the first (0.026–2.63 km), second (2.63–5.24 km), third (5.24–7.85 km) and fourth (7.85–10.46 km) distance classes and (b) a Mantel correlogram between genetic distance and resistance distance showing significant and positive autocorrelation values for the smaller resistance classes. White and black squares represent non-significant and significant relationships between genetic and Euclidean/resistance distances for the different distance classes, respectively.

Landscape variable	Parameter values	Partial Mantel $r$	$P$ -value
Aspect	$90^\circ$ ; $x = 10$ ; $R_{\max} = 2$	<b>0.116</b>	<b>0.024</b>
Land cover	$x = 1$ ; $R_{\max} = 1\,000$	<b>0.112</b>	<b>0.028</b>
Topographic complexity (TC)	$x = 10$ ; $R_{\max} = 2$	<b>0.101</b>	<b>0.042</b>
Rivers	Classified; $R_{\max} = 5\,000$	0.036	0.254
Roads	Classified; $R_{\max} = 2$	0.030	0.289

**Table 1.** The best univariate models of effective landscape resistances based on partial Mantel correlation after removing the effect of the isolation-by-resistance (IBR) model. Models are ranked according to the Partial Mantel  $r$ -value. Optimised parameter values, partial Mantel  $r$  and significance of support are shown. Supported models are indicated in bold.

IBR model (Table 1). Land cover and topographic complexity followed aspect respectively ( $r = 0.112$ ,  $P = 0.028$ ;  $r = 0.101$ ,  $P = 0.042$ ; Table 1). Partial Mantel correlations for natural barriers (rivers) and man-made barriers

Landscape variable	Parameter values	RS <sub>IBR</sub>	(A) <i>r</i>	(A) <i>P</i>	(B) <i>r</i>	(B) <i>P</i>	Supported
Aspect	90°; <i>x</i> = 10; <i>R</i> <sub>max</sub> = 2	<b>0.168</b>	<b>0.116</b>	<b>0.023</b>	<b>-0.052</b>	<b>0.832</b>	<b>Yes</b>
Topographic complexity	<i>x</i> = 10; <i>R</i> <sub>max</sub> = 10	<b>0.094</b>	<b>0.097</b>	<b>0.050</b>	<b>0.003</b>	<b>0.500</b>	<b>Yes</b>
Land cover	<i>x</i> = 0.5; <i>R</i> <sub>max</sub> = 1 000	-0.012	0.112	0.029	0.124	0.0001	No
Rivers	Classified; <i>R</i> <sub>max</sub> = 5	-0.024	0.032	0.278	0.056	0.139	No
Roads	Classified; <i>R</i> <sub>max</sub> = 2	0.021	0.030	0.291	0.009	0.440	No

**Table 2. The best univariate models based on relative support (RS) and causal modelling after removing the effect of the isolation-by-resistance (IBR) model.** Models are ranked with the best-supported model at the top. Optimised parameter values, RS as compared to IBR, partial Mantel *r* and significance of support are shown. Optimised values include equation parameters for *x* (contrast) and *R*<sub>max</sub> (magnitude of the relationship). (A) GD~LV|IBR - partial Mantel test between genetic distance and the landscape variable, partialling out the effect of IBR; (B) GD~IBR|LV - partial Mantel test between genetic distance and IBR distance, removing the effect of the landscape variable. The first column of each test indicates the Mantel *r*-value and the second column the related *P*-value. Supported models are indicated in bold.

	Model	Parameters	RS <sub>IBR</sub>	(A) <i>r</i>	(A) <i>P</i>	(B) <i>r</i>	(B) <i>P</i>	(C) <i>r</i>	(C) <i>P</i>	(D) <i>r</i>	(D) <i>P</i>
1)	A + L + TC	A: 225°; <i>x</i> = 4; <i>R</i> <sub>max</sub> = 1 000	0.053	0.127	0.016	0.074	0.023	A: 0.124	0.004	A: -0.059	0.891
		L: <i>x</i> = 2; <i>R</i> <sub>max</sub> = 1 000						L: 0.086	0.071	L: 0.096	0.014
		TC: <i>x</i> = 4; <i>R</i> <sub>max</sub> = 1 000						TC: 0.110	0.001	TC: -0.073	0.970
2)	A + L	A: 225°; <i>x</i> = 4; <i>R</i> <sub>max</sub> = 1 000	0.018	0.124	0.022	0.106	0.0008	A: 0.144	0.002	A: -0.087	0.966
		L: <i>x</i> = 2; <i>R</i> <sub>max</sub> = 1 000						L: 0.102	0.042	L: 0.081	0.064
3)	A + L	A: 225°; <i>x</i> = 10; <i>R</i> <sub>max</sub> = 2	-0.033	0.115	0.029	0.148	0.0001	-	-	-	-
		L: <i>x</i> = 1; <i>R</i> <sub>max</sub> = 1 000						-	-	-	-
4)	A + TC	A: 225°; <i>x</i> = 4; <i>R</i> <sub>max</sub> = 1 000	0.110	0.113	0.026	0.003	0.485	A: 0.113	0.0001	A: -0.018	0.630
		TC: <i>x</i> = 4; <i>R</i> <sub>max</sub> = 1 000						TC: 0.176	0.0001	TC: 0.050	0.187

**Table 3. The best multivariate models based on relative support (RS), causal modelling after removing the effect of the isolation-by-resistance (IBR) model (A,B) and causal modelling criteria with the reduced model (C,D).** Optimised parameter values, RS as compared to IBR, partial Mantel *r* and significance of support are shown. Optimised values include equation parameters for *x* (contrast) and *R*<sub>max</sub> (magnitude of the relationship). (A) GD~LV|IBR - partial Mantel test between genetic distance and the landscape variable, partialling out the effect of IBR; (B) GD~IBR|LV - partial Mantel test between genetic distance and IBR distance, removing the effect of the landscape variable, (C) GD~LM| - partial Mantel test between genetic distance and the landscape model after removing the effect of the reduced model; (D) G~|LM - partial Mantel test between genetic distance and the reduced model, partialling out the effect of the landscape model. The first column of each test indicates the Mantel *r*-value and the second column the related *P*-value. Model abbreviations: A = Aspect; L = Land cover and TC = Topographic complexity.

(roads) were not significant, regardless of scale and could therefore be excluded from the remaining analyses (Table 1).

Likewise, when variables were evaluated based on relative support (RS) when compared to IBR and the first step of the causal modelling criteria, the best-supported model was aspect, followed by TC (Table 2). Although roads revealed a positive RS value, this model failed (did not pass) the first step of the causal modelling criteria with IBR. In contrast, land cover and rivers showed negative RS values and did not pass the causal modelling and were therefore not supported (Table 2). Parameter values for aspect were the same as the optimisation based on partial Mantel correlation. The *R*<sub>max</sub> parameter value for TC was greater than optimisation based on the partial Mantel correlation. Although land cover did not pass the first step of the causal modelling criteria with IBR and RS, this variable was supported based on the partial Mantel correlation (Table 1) and was therefore included in the multivariate analysis. We concluded that the best univariate models were aspect, land cover and TC although only aspect and TC performed significantly better than distance alone.

Model	Type of model	Variables	$R^2_\beta$	VIF	AICc	$\Delta$ AICc	Weight ( $w_i$ )
A	Reduced	Aspect	0.046	3.12	-12577.60	<b>0.00</b>	0.88
B		Aspect	0.052	3.12	-12573.60	<b>4.03</b>	0.11
		Land cover		1.69			
C		Aspect	0.047	3.12	-12565.70	11.86	0.01
		Rivers		1.54			
D	Full	Aspect	0.051	3.12	-12563.00	14.59	0.00
		Land cover		1.69			
		Rivers		1.54			

**Table 4. Mixed effect models showing the relationship between pairwise genetic distances and resistance distances for different environmental variables.** In order to minimise colinearity among predictors, all variables with VIF values  $> 5$  were removed. VIF = Variance Inflation Factor. The best fitting model was selected using the corrected Akaike Information Criterion (AICc,  $\Delta$ AICc,  $w_i$ ). We used  $R^2$  statistics ( $R^2_\beta$ ) to describe the amount of variation explained by the model. Models with the highest AICc support are in bold ( $\Delta$ AICc  $\leq 2$ ). Marginally supported models are also indicated ( $\Delta$ AICc  $\leq 10$ ).

The best multivariate model based on optimisation with partial Mantel correlation after partialling out the IBR model (see columns A, Table 3) included three variables: aspect, land cover and TC (model 1). Model 2 and 3 were also supported based on the partial Mantel correlation after removing the effect of IBR (Table 3). Model 2 (aspect and land cover), with medium contrast ( $x$ ) and highest resistance ( $R_{\max}$ ) for aspect and medium contrast ( $x$ ) for land cover, was slightly better supported than model 3 with the same variables.

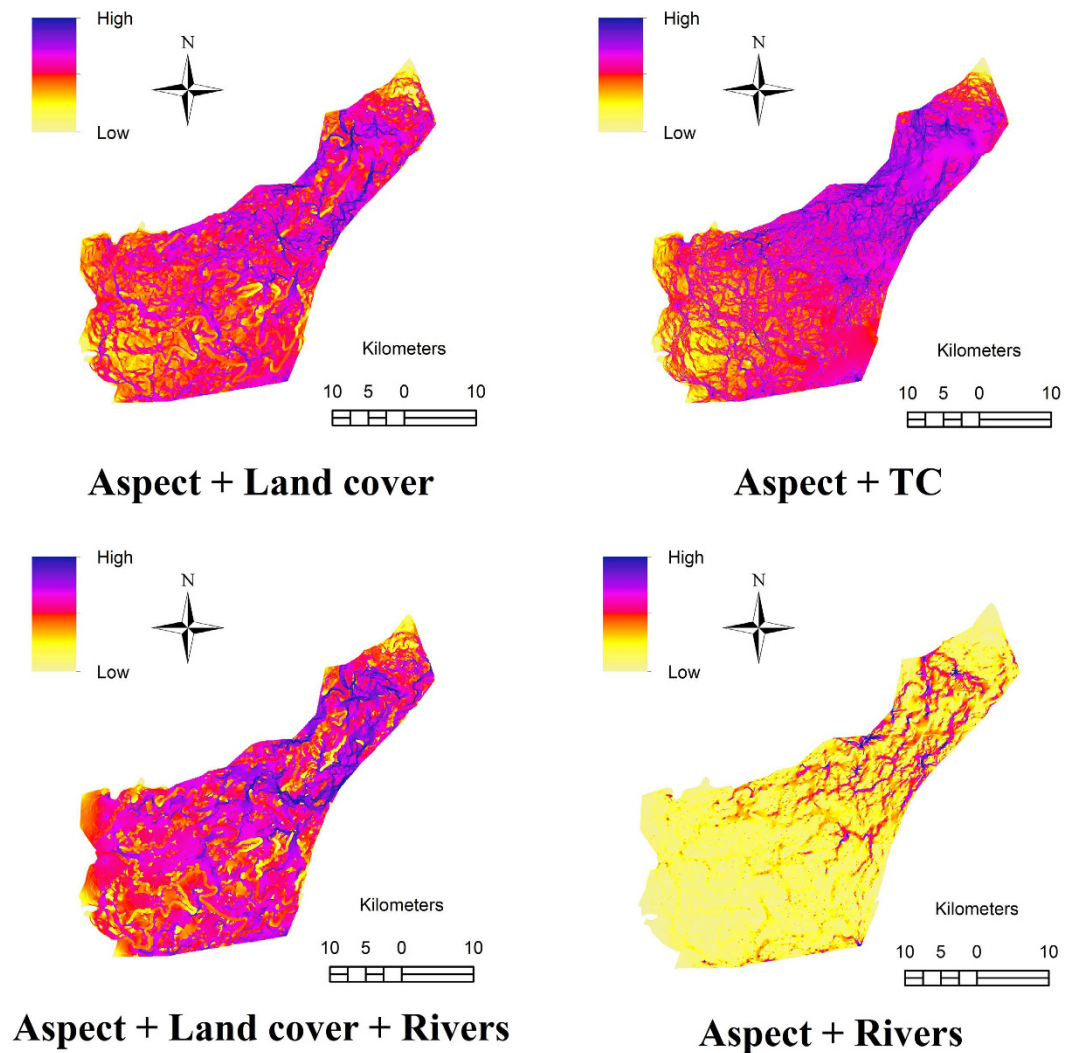
When model optimisation was based on RS, the best multivariate model comprised aspect, land cover and TC (Table 3, model 1). The RS value improved from 0.018 to 0.053 when TC was included. Model 2 also passed RS but model 3 showed a negative value for RS and was therefore not supported. None of the former models passed the causal modelling, which indicated that none of these models was supported independently of the IBR model (see columns B, Table 3, models 1–3).

In addition, when we applied the causal modelling criteria with the reduced model to model 1, the partial Mantel correlation was significant ( $P = 0.001$ ) when partialling out the effect of the reduced model ( $GD \sim (A + L + TC) | (A + L)$ ), while the opposite ( $GD \sim (A + L) | (A + L + TC)$ ) was not significant ( $P = 0.970$ ), indicating that the inclusion of TC improved the model (Table 3). Likewise, aspect passed the causal modelling with the reduced model. However, land cover failed the causal modelling with the reduced model so the only variables that were included in the final multivariate model were aspect and TC (Table 3, model 4). Model 2 also passed the causal modelling with the reduced model. The only supported models in the multivariate framework were  $A + L$  and  $A + TC$  (Table 3; model 2 and 4). When we compared models 2 and 4 against each other and IBR using RS, model  $A + TC$  performed best (see Supplementary Table S3). The RS for model  $A + TC$  compared to  $A + L$  and IBR was only slightly greater than zero, but it still passed the causal modelling criteria for significance (Table 3). In addition, model  $A + TC$  performed significantly better ( $r = 0.113$ ,  $P = 0.026$ ) than distance alone ( $r = 0.003$ ,  $P = 0.485$ ). We concluded that the best multivariate model was  $A + TC$ , indicating that aspect influenced gene flow and this pattern is driven by the availability of water and favourable vegetation. Topographic complexity was also likely to influence gene flow.

**Mixed effect models.** Statistical models based on resistance distances suggested that environmental variables explained more variation in genetic differentiation than the IBR model that only explained approximately 3% of the variance in genetic distance. The following variables were highly correlated (VIF values  $> 5$ ) and thus excluded from the linear mixed effect models: TC, IBR, roads and geographic distance. We therefore built a full model (aspect, land cover and rivers) in an attempt to exclude correlated variables. After the inclusion of the remaining landscape variables the amount of variation that has been explained increased to approximately 5% (Table 4). The best-supported model based on the corrected Akaike Information Criterion (AICc; Table 4, model A) included aspect alone (AICc = -12577.60;  $w_i = 0.88$ ). In general, models within two AIC units of the best-supported model are interchangeable and models with a  $AICc \leq \sim 10$  show marginal support. We therefore inferred that the reduced model B (Table 4) that included aspect and land cover (AICc = -12573.60;  $w_i = 0.11$ ) and model C (aspect and rivers) showed marginal support.

Although model B only showed marginal support based on AICc, this model explained more of the variation (5.2%) in the dependent variable followed by model D that explained 5.1% of the variation (Table 4). Model D comprised aspect, land cover and rivers. In addition, the  $R^2_\beta$  values of model D (full model) and model B were similar ( $R^2_\beta = 0.051$ ). Likewise, models A (aspect;  $R^2_\beta = 0.046$ ) and C (aspect + rivers;  $R^2_\beta = 0.047$ ) had similar  $R^2_\beta$  values that explained slightly less of the variation in genetic distance. Based on mixed effect models, we concluded that aspect, land cover and rivers were the variables that explained variation best.

Based on our best-supported multivariate hypotheses, the most likely regions of genetic connectivity for *M. natalensis* are represented in Fig. 3. Areas with highest resistance to gene flow are indicated in light yellow. All maps showed high spatial heterogeneity in connectivity with a north to south gradient of high to low connectivity except for map “Aspect + Rivers”. Eastern aspect and land cover presented the least resistance to dispersal/movement through the landscape for this small mammal. Map “Aspect + Rivers” showed that the rivers in the



**Figure 3.** Current maps generated in CIRCUITSCAPE showing connectivity between 101 *Mastomys natalensis* trapping transects from Hluhluwe-iMfolozi Park, South Africa for the following best-supported landscape hypotheses: “Aspect + Land cover”, “Aspect + Topographic complexity”, “Aspect + Land cover + Rivers” and “Aspect + Rivers”. Dark blue represents areas with highest current densities whereas areas with highest resistance (lowest current densities) are represented in the light yellow colour. Areas indicated in dark blue will therefore facilitate gene flow (higher connectivity) whereas areas in light yellow may restrict gene flow. Maps were modified in ArcGIS v 10.1. (<http://www.esri.com/software/arcgis/arcgis-for-desktop>).

park exhibit low current density (restriction to gene flow). This map also showed no clear north to south gradient of high to low connectivity but rather displayed high landscape connectivity in the northern part of the park.

### Discussion

We present the first landscape genetics study for an African small mammal, identifying features that may potentially explain contemporary genetic structure in a relatively unmodified landscape. Our results showed a higher and significant genetic correlation among individuals than expected at smaller geographic scales (up to about 11 km) confirming the results of van Hooft *et al.*<sup>3</sup> suggesting that *M. natalensis* exhibit a pattern of kin clustering at smaller geographic scales while, long distance dispersal events result in no pattern of IBD as suggested by spatial autocorrelation at larger geographic scales. It has also been documented that if animals breed at a young age that they will show a greater tendency to breed close to their natal site<sup>14</sup>. This further supports the pattern of IBD as indicated by spatial autocorrelation at smaller geographic scales. Long distance dispersal will erode IBD patterns and therefore weak patterns of IBD could be due to rare long distance dispersal as we report in this study for *M. natalensis* and as has been found for other small mammals<sup>26</sup>.

Low to moderate differentiation indicates a pattern of sub-clustering of up to four potentially relevant clusters that differ in their geographic coherence/genomic integrity that does not coincide with those expected if divergence was driven solely by isolation-by-distance (all  $F_{ST}$  values were significant and ranged from 0.025–0.124 between clusters). The most closely related individuals are not necessarily those that come from the same

geographical area. This indicates that forces imposing population genetic structuring are not based purely on IBD as suggested by spatial autocorrelation. Rather, movement seems to be restricted providing support for the hypothesis that restricted dispersal is one of the forces driving patterns of genetic structure in *M. natalensis*. This has also been evident in other landscape genetic studies of rodent species<sup>27</sup>. Complex and inconsistent patterns of genetic heterogeneity have been described as “chaotic genetic patchiness” consistent with the patterns observed in this study<sup>28</sup> and the landscape features investigated in this study may underlie this pattern.

In the broader sense, a landscape genetics approach should be incorporated into spatial models of disease processes to increase the ability to predict patterns of disease occurrence, to prevent the spread of disease and to develop management policies<sup>29</sup>. Likewise, the identification of barriers to movement in mammal disease reservoirs is important as this can assist in disease containment strategies<sup>30</sup>, pest and crop management<sup>31</sup>.

The results of the univariate/multivariate parameter optimisation suggest that slope aspect is the landscape variable most strongly correlated with gene flow in *M. natalensis*, followed by land cover, topographic complexity (TC) and rivers. Slope aspect promotes gene flow in *M. natalensis*, with western facing slopes posing greater resistance than eastern facing slopes. Similarly, Castillo *et al.*<sup>32</sup> provided evidence that eastern facing slopes are positively correlated with genetic connectivity in the American pika. Land cover is also strongly correlated with gene flow in *M. natalensis* with thicket vegetation as the most favourable. TC and rivers (movement across rivers) are the two landscape features that most strongly restrict gene flow. This suggests that *M. natalensis* dispersal is mainly restricted by physical limitations. Complex topographies such as ridges and steep slopes pose more resistance to movement. Resistance to gene flow increases as a landscape becomes more complex because of greater energetic cost to movement due to the small body size of *M. natalensis*. Rivers/streams may act as barriers to gene flow as is evident from the Bayesian clustering analysis (Cluster I). Although not explicitly tested in this study, the size/width of the watercourse/river is also likely to determine the extent to which a barrier is permeable. Having stated this, there may or may not be facilitation of gene flow along rivers. In general, rivers appear to be barriers to gene flow but it is evident from this study that *M. natalensis* are able to traverse these barriers confirming their swimming behaviour as described by Hickman and Machiné<sup>11</sup>. Watercourses have shown to have a small effect on gene flow and may impede movement but it does not act as a complete barrier to gene flow. Likewise, it has been shown that migration between small mammal sub-populations occurs across rivers<sup>27</sup>.

*Mastomys natalensis* is commonly found across a large range of different habitats except deserts and very high altitudes<sup>24</sup>. Although classified as a generalist, some landscape features may still have an influence on gene flow in this small mammal. A number of studies have suggested that landscape features can be prominent in shaping genetic structure in generalist species. For example, it has been shown that a river and a motorway represented a barrier to dispersal in the Eurasian badger (*Meles meles*)<sup>33</sup>. Genetic differentiation for habitat specialists is expected to be larger than for generalist species<sup>34</sup>. Here, we identify aspect and the associated vegetation cover as important landscape features in shaping the genetic structure in this generalist species.

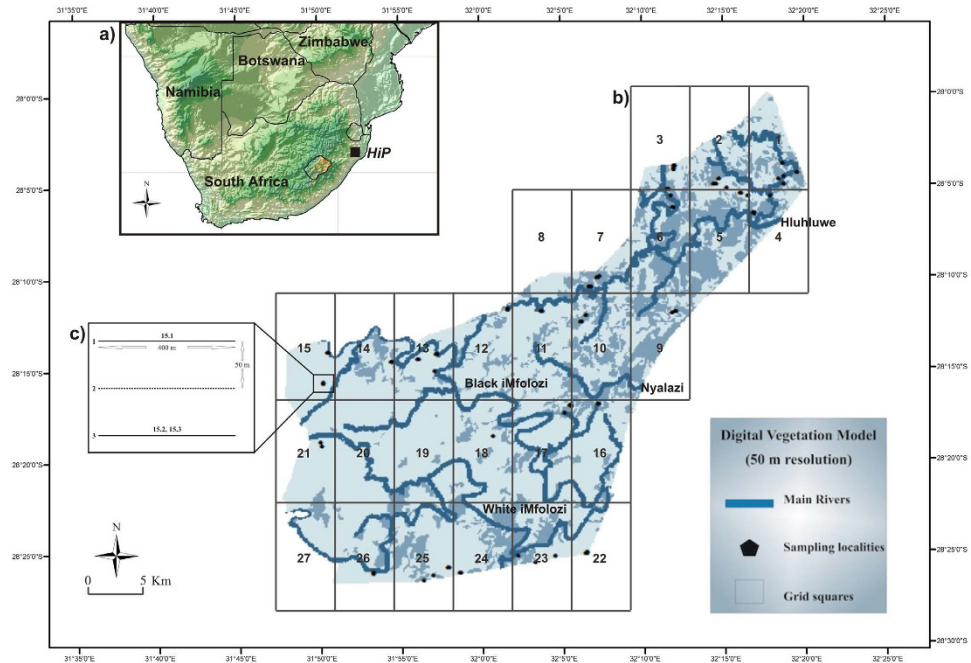
Due to the fact that this park is in the southern hemisphere, western facing slopes represent a heterogeneous dry tropical savanna habitat. In contrast, the eastern facing slopes are characterised by a mesic and cooler habitat (lower solar radiation) with dominant vegetation featuring wooded trees<sup>35</sup>. Eastern facing slopes are in general cooler (lower temperatures because of morning sun) whereas western facing slopes experience higher afternoon temperatures. In addition, southern/eastern facing slopes in the region tend to desiccate more slowly than the northern/western facing slopes (G Clinning, personal communication, Ezemvelo KZN Wildlife).

South Africa is in general characterised by a decrease in rainfall from east to west and the eastern regions of the coastal plateau are more humid<sup>36</sup>. Southern/eastern facing slopes are therefore wetter in general than northern/western facing slopes because moisture and air masses originate from the Indian Ocean (see Supplementary Fig. S5)<sup>36</sup>. These air masses cause the trapping of rain that moves inland<sup>37</sup>. This area of South Africa is also known as the “mist belt”; the mist is usually associated with southerly winds providing moisture to eastern facing slopes<sup>36</sup>. This higher rainfall provides suitable conditions to sustain thicket vegetation however, there may be dry periods that prevent the development of forest<sup>38</sup>. Thicket is renowned for its high biodiversity and includes evergreen, sclerophyllous or succulent trees, shrubs and vines<sup>38</sup>. Thicket often comprises dense impenetrable vegetation (shrubs) with no distinct layer of trees and shrubs<sup>38</sup>, providing small mammals with cover and protection against predators while foraging and dispersing. In addition, thicket may provide a greater variety of food sources such as fruits and vegetative plant material, i.e. leaves, stems and seeds<sup>10</sup>. The apparent preference of *M. natalensis* for eastern facing slopes (areas of higher rainfall) and thicket vegetation (areas of seasonally available water) demonstrate the species’ dependency on seasonally available water<sup>10</sup>.

Dispersal, resource availability and population densities affect genetic structure in natural populations<sup>39</sup> but landscape heterogeneity may also promote/facilitate gene flow between populations. Patterns of genetic sub-structuring observed in this study may be explained by a combination of shorter-distance movements (when population densities are low) with increasing distances when population densities increase<sup>40</sup>.

We used a range of analytical and modelling approaches, all with some limitations. Mantel and partial Mantel tests have been widely used in landscape genetic studies<sup>41,42</sup> although some authors have debated the use of these tests due to the potential increase in Type I error<sup>43</sup>. Error can be avoided by evaluating the correlation coefficient itself as opposed to the test for significance, an approach that may explain why landscape variables such as rivers and land cover are not supported in the present study using the partial Mantel framework. In addition, the effect of rivers on gene flow was not strong enough to be detected in the partial Mantel tests. We therefore implemented mixed effect models on distance matrices since analysing data using only one method could result in false, method-dependent outcomes<sup>43</sup>.

In light of global warming, the dry regions of the subtropics are likely to get drier which will result in less cover and an absence of thicket vegetation. As such, population connectivity of *M. natalensis* may be affected by climate change. By using *M. natalensis* in a landscape genetics context, one can expand the results to the ecosystem management of other small mammal species with similar ecosystem functions. In addition, apart from contributing to



**Figure 4.** Map of the Hluhluwe-iMfolozi Park (HiP) situated in the KwaZulu-Natal province (see map (b)) of South Africa (see map (a) indicated by black square). The park has been divided into 27 grids and samples were collected to represent all habitat types. Sampling coordinates along each transect within a grid were combined into a midpoint coordinate for that transect. Main rivers (in blue) and thicket vegetation (in light blue-grey) are also indicated on the map. Inset (c) shows a schematic diagram of transect layout for grid 15. This map was modified in CorelDraw Graphics Suite X3 (2014; <http://www.coreldraw.com>).

landscape genetic studies in Africa<sup>21</sup> the present study may have implications in epidemiological and agricultural research associated with problem rodents.

## Methods

**Study site, sampling and genotyping.** Hluhluwe-iMfolozi Park (HiP) is situated in the northern escarpment of the KwaZulu-Natal province of South Africa, covering an area of approximately 960 km<sup>2</sup> (28°17'49"S; 31°44'32"E) and contains four large rivers: the Hluhluwe, Nyalazi, Black and White iMfolozi Rivers (Fig. 4)<sup>44</sup>. The climate is coastally modified and varies according to topography with a mean annual rainfall of 985 mm at higher altitude<sup>44</sup>. Annual temperatures range from 13 °C to 35 °C and the park is situated in the Savanna biome of South Africa, supporting a variety of habitat types from scrap forest to thicket<sup>38</sup>.

Rodent samples were collected using a landscape grid (see Supplementary Table S1). All permit information was reported (see Supplementary Table S2). The sampling method was carried out in accordance with Ezemvelo KZN Wildlife protocols and animals were handled under the guidelines of the American Society of Mammalogists (ASM; Animal Care and Use Committee, 2011)<sup>45</sup>. Ear biopsy sampling received ethical approval from Cardiff University in September 2010.

The park was divided into 27 sections, each being approximately 10 km North-South × 5 km East-West (Fig. 4). One of the sections (8) was not sampled since it only included a small part of the park and was inaccessible (Fig. 4). No *M. natalensis* could be captured from three (section 19, 20 and 27) of the remaining 26 sections. We therefore sampled 23 sections across the park. A total of 150 Sherman traps (H.B. Sherman Traps Inc. Florida, USA) were placed in each section under logs, rocks and tussocks in an attempt to stratify sampling across the landscape. Sampling was carried out up to a maximum of four nights per section, by using three transects per night approximately 50 m apart (~400 (l) × 100 (w) m), with a 10–15 m inter-trap interval (Fig. 4, inset c). Up to 21 individuals were collected per section over a total of 4,516 trap nights, resulting in a total of 272 samples, 260 of which were analysed; twelve being excluded because of poor amplification success rate. Tissue samples for DNA analysis were collected by ear biopsy and stored in 99% ethanol at –20 °C.

Total genomic DNA was extracted from tissue using a DNeasy<sup>®</sup> Tissue Kit (QIAGEN<sup>®</sup> Hilden, Germany) following the manufacturer's instructions. Genotyping was carried out using 10 microsatellite markers originally isolated from *Mastomys huberti*<sup>46,47</sup>. We followed PCR conditions as described in Galen *et al.*<sup>46</sup> and Loiseau *et al.*<sup>47</sup>. Amplifications were carried out in a total volume of 10 µl containing 5 µl of QIAGEN Multiplex PCR Master Mix, 0.2 µM of each of the forward and reverse primers and 1 µl of DNA (~50 to 100 ng/µl). The ten primer pairs were divided into three multiplexes and one singleplex using three dyes (FAM, HEX and TAM): MH28, MH30, MH60, MH133, MH141 (Multiplex 1); MH05, MH74 (Multiplex 2); MH105, MH188 (Multiplex 3) and MH80 (Singleplex 1). PCR products were processed commercially by Macrogen Inc, Korea ([www.macrogen.com/eng/](http://www.macrogen.com/eng/)).

Electropherograms were scored using GeneMarker v 1.91 (SoftGenetics LLC). Observed allele lengths were binned where necessary and rounded to integers using the software TANDEM v 1.8<sup>48</sup>.

**Statistical analyses.** We checked for the presence of null alleles using MICROCHECKER<sup>49</sup>. The rarefaction procedure implemented in FSTAT v 2.9.3.2<sup>50</sup> was used to estimate the expected number of alleles and to compare allelic richness ( $r$ ). Hardy-Weinberg equilibrium (HWE) was analysed for each locus using FSTAT v 2.9.3.2<sup>50</sup>. Linkage disequilibrium (LD) between all pairs of loci and  $F$ -statistics between clusters were estimated using GENEPOP v 4.1<sup>51</sup>. Statistical significance for HWE and LD was assessed by  $P$ -values and the implementation of the modified false discovery rate (FDR) method<sup>52</sup> for multiple tests. Observed ( $H_O$ ) and unbiased expected ( ${}_U H_E$ ) heterozygosities were estimated for all loci using GENETIX v 4.5.0.2<sup>53</sup>.

**Population structure.** Bayesian clustering in STRUCTURE v 2.2.3<sup>54</sup> was used to infer the most likely number of population clusters ( $K$ ). The analysis was implemented for 450 000 iterations following a burn-in of 50 000 iterations with no *a priori* locality data. Both the posterior probability of the data for the given value of  $K$  ( $\text{LnPr}(X|K)$ ) and its rate of change ( $\Delta K$ )<sup>25</sup> were used to evaluate population structure. Twenty independent runs were carried out for  $K$  values from one to ten and an admixture model with correlated allele frequencies<sup>54</sup> was assumed. Progressive partitioning at various hierarchical levels using  $K = 2$ <sup>55</sup> was used to further examine structure within resolved clusters.

**Genetic distance and spatial autocorrelation.** We analysed the genetic data (a reduced dataset of 216 individuals) using transect lines as the unit of spatial analysis. The geographical midpoints of each transect were used as location coordinates (see Supplementary Table S1). We estimated pairwise genetic distance for each transect by using the proportion of shared alleles in MSA v 4.05<sup>56</sup> (Table S4). Euclidean geographic distances in kilometres were calculated between all transects and pairwise resistance distances were estimated using CIRCUITSCAPE v 3.5<sup>57</sup> by creating a raster file with a constant value of 1 for all pixels (isolation-by-resistance (IBR) model).

A Mantel correlogram using the per transect association between the proportion of shared alleles and geographic or resistance distances ( $r_M$ ) was calculated to test for spatial autocorrelation using a permutation test of significance (10 000 iterations). The modified false discovery rate method<sup>52</sup> was employed to correct for global significance.

**Landscape resistance optimisation (univariate and multivariate models).** Landscape resistance was modelled as a function of aspect, river, roads, topographic complexity (TC) and land cover according to a landscape resistance hypothesis (see Table 5 for summary). Optimised values included equation parameters for  $x$  (contrast) and  $R_{\max}$  (magnitude of the relationship). For a full description see the Supplementary information. Pairwise landscape resistance matrices for each resistance surface were estimated using CIRCUITSCAPE v 3.5<sup>57</sup> by implementing the pairwise mode option with focal points, connecting eight neighbours based on the average resistance.

Partial Mantel tests have been widely used in landscape genetic studies although some authors have debated the use of these tests<sup>58</sup>. Partial Mantel tests can be successful when used in the causal modelling framework as was done in this study<sup>41</sup>.

Genetic distance was compared with landscape resistances to identify the functional scale for which each variable best explained the observed genetic structure. Scaled transformations for each landscape variable based on a unimodal peak of support<sup>59</sup> in the partial Mantel correlation coefficient were used to rank variables/models. All statistical analyses were performed in R<sup>60</sup>.

**Univariate and multivariate model selection.** We used a combination of causal modelling approaches proposed by Cushman *et al.*<sup>58</sup> and Wasserman *et al.*<sup>61</sup>. The first step<sup>58,61</sup> suggests that if a resistance hypothesis is supported independently of the null model then: (A) partial Mantel tests between the genetic distance and landscape variable would be significant after removing the effect of IBR (GD~LV|IBR); (B) partial Mantel tests between genetic distance and IBR distance would not be significant, partialling out the landscape variable (GD~IBR|LV). The second step<sup>58</sup> allows for the comparison of causal modelling with a reduced model. If a landscape model (true model) is supported independently of the other candidate models then: (C) partial Mantel tests between genetic distance and the landscape model would be significant, removing the effect of the reduced model (GD~LM); (D) partial Mantel tests between genetic distance and the reduced model would not be significant, partialling out the effect of the landscape model (GD~|LM).

Candidate models with similar parameters to the top model were assessed against each other rather than simply against IBR to evaluate models based on relative support (RS). Relative support can be defined as:  $RS_{A|B} = (GD \sim A|B) - (GD \sim B|A)$  where GD is the genetic distance, A is the resistance distance matrix for variable A, B is the resistance matrix obtained for variable B and (GD~A|B) is the partial Mantel  $r$  between GD and variable A after removing the effect of variable B. The landscape model with a positive RS in all comparisons represents the best model<sup>58</sup>. In order for landscape models to be accepted, models also had to pass the causal modelling criteria with IBR.

Landscapes are complex structures and therefore best described using multivariate approaches that incorporate multiple landscape features. Here, we built rasters equal to the sum of the univariate rasters for each landscape variable. The two landscape variables with the highest partial Mantel  $r$  when removing the effect of IBR were used to create a series of bivariate models by keeping the parameters constant for the first variable, while varying the parameters for the second variable. The best-supported model for the second variable was identified based on the partial Mantel correlation removing the effect of IBR. Subsequently, we identified the optimum parameters for the first variable by holding the second constant. Once we identified the optimum parameters for

Variable	Hypothesis	Parameters	Maps	
Aspect	The optimal aspect (less resistance) is associated with the availability of water and favorable vegetation	$x = 0.5, 1, 2, 4, 10$	200	
		$R_{\max} = 2, 10, 100, 500, 1\ 000$		
		$\theta_{opt} = 0, 45, 90, 135, 180, 225, 275, 315, 360$		
		$\theta = 0, 45, 90, 135, 180, 225, 275, 315, 360$		
		Flat areas = $R_{\max}/2$		
Rivers	Physical barrier to small mammal movement	$R_{\max} = 2, 5, 10, 50, 100, 250, 500, 750, 1\ 000, 5\ 000, 10\ 000$	22	
		Land = 1		
Roads	Physical barrier to small mammal movement	$R_{\max} = 2, 5, 10, 50, 100, 250, 500, 750, 1\ 000, 5\ 000, 10\ 000$	22	
		Land = 1		
TC	Resistance to gene flow increases as landscape becomes more complex	$x = 0.5, 1, 2, 4, 10$	150	
		$R_{\max} = 2, 10, 100, 500, 1\ 000$		
		Radii = 1, 2, 5, 10, 25, 50		
Land cover	Land cover as a source of food and cover against predators promotes gene flow	Array of resistance values (permutations) = 1, 168, 334, 501, 667, 1 000	120	
		$R_{FH} = 1; R_{\max} = 2, 5, 10, 50, 100, 250, 500, 750, 1\ 000, 5\ 000, 10\ 000, 100\ 000$		36
		$x = 0.5, 1, 2, 4, 10; R_{\min} = 1; R_{\max} = 2, 5, 10, 50, 100, 250, 500, 750, 1\ 000$		180

**Table 5. A summary of the landscape variables and the corresponding resistance hypotheses.** Parameter values for each variable are indicated. These include  $x$  (power function),  $R_{\min}/R_{\max}$  (minimum or maximum resistance),  $\theta_{opt}$  (hypothesised optimal aspect in increments of  $45^\circ$  from  $0^\circ$  to  $315^\circ$ ),  $\theta$  (aspect value in increments of  $45^\circ$  from  $0^\circ$  to  $315^\circ$ ), radii (buffer area in number of cells) and  $R_{FH}$  (favourable habitat). Abbreviations: TC = Topographic complexity. The total number of maps generated for each variable are indicated in the last column. For a full description see the Supplementary information.

the first two models, additional landscape variables were added while keeping the first two model parameters constant. We re-optimised the remaining variables until the best-supported model did not change. All models were also evaluated using their RS by varying the model parameters for one variable while holding the others constant until variable parameters stabilised. Models were required to pass the two causal modelling steps and variables were also required to pass the causal modelling criteria with the reduced model. For example, in order for variable C to be accepted the following needs to be true for the different multivariate rasters:  $GD \sim (A + B + C) | (A + B)$  must be significant and  $GD \sim (A + B) | (A + B + C)$  must be non-significant. Multivariate rasters  $(A + B + C; A + B)$  contained the combined information of variables A, B and C. Partial Mantel tests were performed in R using the “vegan” package<sup>62</sup>.

**Mixed effect models.** In addition to causal modelling, we also implemented linear mixed effect models in R using the “lme4” package<sup>63</sup> to account for dependency between pairwise observations in a distance matrix<sup>64</sup>. To correct for the dependency among pairwise data points in our data, we followed the maximum likelihood populations-effects (MLPE) method as described in Clarke *et al.*<sup>65</sup> and more recently in Van Strien *et al.*<sup>66</sup>. Sampling unit was introduced as a random effect term that accounted for the non-independency of pairwise sampling unit distances by assuming an intercept that was different for each unit. The random effect term in the model specified the pairwise sampling unit structure of the dataset. In addition, all explanatory variables were introduced as fixed effect terms. Restricted maximum likelihood (REML) estimates of the intercept were the same as for those obtained from linear regression<sup>65</sup> because all explanatory variables were standardised. In each model, the random effect was standard with different explanatory variables. First we calculated the Variance Inflation Factors (VIF) for each variable in the model<sup>67</sup>. Variance Inflation Factors values above 5 show evidence for collinearity and explanatory variables exhibiting significant levels of multi-collinearity were removed from the models to minimise potential error<sup>68</sup>.

We built a full model with all variables to identify the combined effects of multiple variables on gene flow. This model only included the variables that did not show any evidence of collinearity. The full model was refined by using multi-model inference in the “MuMIN” package<sup>69</sup> of R. Maximum-likelihood populations-effects (MLPE) models were fitted with REML estimation to account for the association of each pairwise distance between sampling units. This method is desirable for unbiased estimates of the variance components of mixed models<sup>65</sup>. We used model averaging from a global model to select the best models by calculating the corrected Akaike Information Criterion values (AICc)<sup>70</sup>. Akaike Information Criterion scores are often represented as  $\Delta AICc$  scores, which indicates the difference between the best model ( $\Delta AICc = 0$ ) and all the other candidate models. We also calculated AICc weights ( $w_i$ ). Models with the lowest change in the AICc score ( $\Delta AICc = 0$ ) and the highest Akaike weight were considered the best<sup>71</sup>. Models within two AIC units of the best-supported model were interchangeable and models with a  $\Delta AICc \leq 10$  showed marginal support<sup>70</sup>.

In addition, we calculated the  $R^2_\beta$  statistic<sup>72</sup>, which compared a model with fixed effects to a null model (we used the IBR model as the null model). The null model comprised of the random effect and an intercept. The  $R^2_\beta$  was calculated from the Kenward-Rodger  $F$  and degrees of freedom<sup>73</sup>. We used the “KRmodcomp” function from the R package “pbkrtest”<sup>74</sup>. We used the  $R^2_\beta$  statistic to describe the proportion of variation being explained by the models.

**Current maps.** Current maps based on our best-supported multivariate hypotheses were generated in CIRCUITSCAPE v 3.5<sup>57</sup> in order to indicate the most likely regions of genetic connectivity/movement for *M. natalensis* within the Hluhluwe-iMfolozi Park.

**Data accessibility.** Microsatellite data and GIS raster maps available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.492b8>.

## References

- Jones, K. E. & Safi, K. Ecology and evolution of mammalian biodiversity. *Philos. T. Roy. Soc. Lon. B* **366**, 2451–2461 (2011).
- Pearce, J. & Venier, L. Small mammals as bioindicators of sustainable boreal forest management. *For. Ecol. Manag.* **208**, 153–175 (2005).
- Van Hooft, P., Cosson, J. F., Vibe-Petersen, S. & Leirs, H. Dispersal in *Mastomys natalensis* mice: use of fine-scale genetic analyses for pest management. *Hereditas* **145**, 262–273 (2008).
- Luis, A. D. *et al.* A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc. Roy. Soc. Lon. B* **280**, 20122753 (2013).
- Margules, C. R. & Pressey, R. L. Systematic conservation planning. *Nature* **405**, 243–253 (2000).
- Le Galliard, J. F., Rémy, A., Ims, R. A. & Lambin, X. Patterns and processes of dispersal behaviour in arvicoline rodents. *Mol. Ecol.* **21**, 505–523 (2012).
- Estes-Zumpf, W. A., Rachlow, J. L., Waits, L. P. & Warheit, K. I. Dispersal, gene flow, and population genetic structure in the pygmy rabbit (*Brachylagus idahoensis*). *J. Mammal.* **91**, 208–219 (2010).
- Patton, J. L., Da Silva, M. N. F. & Malcolm, J. R. Hierarchical genetic structure and gene flow in three sympatric species of Amazonian rodents. *Mol. Ecol.* **5**, 229–238 (1996).
- Bohonak, A. J. Dispersal, gene flow and population structure. *Q. Rev. of Biol.* **74**, 21–45 (1999).
- Chimimba, C. T. & Bennett, N. Order: Rodentia in *The Mammals of the Southern African Subregion 3<sup>rd</sup> edn* (eds Skinner, J. D. *et al.*) Subfamily Murinae, 127–162 (Cambridge University Press, 2005).
- Hickman, G. C. & Machiné, C. Swimming behaviour in six species of African rodents (Cricetidae, Muridae). *Acta Theriol.* **31**, 449–466 (1986).
- Monath, T. P., Newhouse, V. F., Kemp G. E., Setzer, H. W. & Cacciapuoti A. Lassa virus isolation from *Mastomys natalensis* rodents during an epidemic in Sierra Leone. *Science* **185**, 263–265 (1974).
- Lecompte, E., Fichet-Calvet, E. & Daffis, S. *Mastomys natalensis* and Lassa Fever, West Africa. *Emerg. Infect. Dis.* **12**, 1971–1974 (2006).
- Leirs, H., Verheyen, W. & Verhagen, R. Spatial patterns in *Mastomys natalensis* in Tanzania (Rodentia: Muridae). *Mammalia* **60**, 545–555 (1996).
- Brouat, C., Loiseau, A., Kane, M., Bâ, K. & Duplantier, J. M. Population genetic structure of two ecologically distinct multimammate rats: the commensal *Mastomys natalensis* and the wild *Mastomys erythroleucus* in southeastern Senegal. *Mol. Ecol.* **16**, 2985–2997 (2007).
- Manel, S., Schwartz, M. K., Luikart, G. & Taberlet P. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* **18**, 189–197 (2003).
- Gaufre, B., Estoup, A., Bretagnolle, V. & Cosson F. Spatial genetic structure of a small rodent in a heterogeneous landscape. *Mol. Ecol.* **17**, 4619–4629 (2008).
- Coulon, A., Guillot, G. & Cosson, J. F. Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Mol. Ecol.* **15**, 1669–1679 (2006).
- Meyrom, K. Y. *et al.* Nest-box use by the Barn Owl *Tyto alba* in a biological pest control program in the Beit She’an valley, Israel. *Ardea* **97**, 463–467 (2009).
- Taylor, Z. S. & Hoffman, S. M. G. Landscape models for nuclear genetic diversity and gene structure in white-footed mice (*Peromyscus leucopus*). *Heredity* **112**, 588–595 (2014).
- Storfer, A., Murphy, M. A., Spear, S. E., Holderegger, R. & Waits L. P. Landscape genetics: where are we now? *Mol. Ecol.* **19**, 3496–3514 (2010).
- Delaney, K. S., Riley, S. P. D. & Fisher, R. N. A rapid, strong, and convergent genetic response to urban habitat fragmentation in four divergent and widespread vertebrates. *PLoS ONE* **5**, e12767 (2010).
- Schwenk, W. S. & Donovan, T. M. A multispecies framework for landscape conservation planning. *Conserv. Biol.* **25**, 1010–1021 (2011).
- Colangelo, P. *et al.* A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. *Biol. J. Linn. Soc.* **108**, 901–916 (2013).
- Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
- Munshi-South, J. Urban landscape genetics: canopy cover predicts gene flow between white-footed mouse (*Peromyscus leucopus*) populations in New York City. *Mol. Ecol.* **21**, 1360–1378 (2012).
- Vignieri, S. N. Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse (*Zapus trinotatus*). *Mol. Ecol.* **14**, 1925–1937 (2005).
- Planes, S. & Lenfant, P. Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. *Mol. Ecol.* **11**, 1515–1524 (2002).
- Real, L. A. & Biek, R. Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. *J. R. Soc. Interf.* **4**, 935–948 (2007).
- Côté, H., Garant, D., Robert, K., Mainguy, J. & Pelletier, F. Genetic structure and rabies spread potential in raccoons: the role of landscape barriers and sex-biased dispersal. *Evol. Appl.* **5**, 393–404 (2012).
- Sluydts, V., Davis, S., Mercelis, S. & Leirs, H. Comparison of multimammate mouse (*Mastomys natalensis*) demography in monoculture and mosaic agricultural habitat: Implications for pest management. *Crop. Prot.* **28**, 647–654 (2009).
- Castillo, J. A., Epps, C. W., Davis, A. R. & Cushman, S. A. Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Mol. Ecol.* **23**, 843–856 (2014).
- Frantz, A. C., Pope, L. C. & Etherington, T. R. Using isolation-by-distance-based approaches to assess the barrier effect of linear landscape elements on badger (*Meles meles*) dispersal. *Mol. Ecol.* **19**, 1663–1674 (2010).
- Geffen, E., Anderson, M. J. & Wayne, R. K. Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Mol. Ecol.* **13**, 2481–2490 (2004).
- Pooley, E. *The Complete Field Guide to Trees of Natal, Zululand and Transkei 1<sup>st</sup> edn* (Durban: Natal Flora Publications Trust, Pietermaritzburg, 1993).

36. Jackson, S. P. Climates of southern Africa. *S. Afr. Geogr. J.* **33**, 17–37 (1951).
37. Bruen, M. Hydrological considerations for development in the East Usambara Mountains in *Forest Conservation in the East Usambara Mountains, Tanzania* (eds Hamilton, A. et al.) Ch. 15, 117–140 (IUCN, Gland Switzerland and Cambridge, UK, 1989).
38. Rutherford, M. C., Mucina, L. & Powrie, L. W. Biomes and bioregions of southern Africa in *The Vegetation of South Africa, Lesotho and Swaziland* (eds Mucina, L. et al.) Ch. 3, 32–51 (Strelitzia 19, South African National Biodiversity Institute, Pretoria, 2006).
39. Peacock, M. M. & Ray, C. Dispersal in pikas (*Ochotona princeps*): combining genetic and demographic approaches to reveal spatial and temporal patterns in *Dispersal* (eds Clobert, J. et al.) Ch. 4, 43–56 (New York, Oxford University Press, 2001).
40. Monadjem, A. & Perrin, M. R. The effect of supplementary food on the home range of the multimammate mouse *Mastomys natalensis*. *S. Afr. J. Wildl. Res.* **28**, 1–3 (1998).
41. Cushman, S. A. & Landguth, E. L. Spurious correlations and inference in landscape genetics. *Mol. Ecol.* **19**, 3592–3602 (2010).
42. Marrotte, R. R., Gonzalez, A. & Millien, V. Landscape resistance and habitat combine to provide an optimal model of genetic structure and connectivity at the range margin of a small mammal. *Mol. Ecol.* **23**, 3983–3998 (2014).
43. Balkenhol, N., Waits, L. P. & Dezzani, R. J. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography* **32**, 818–830 (2009).
44. Whateley, A. & Porter, R. N. The woody vegetation communities of the Hluhluwe-Corridor-Umfolozi Game Reserve Complex. *Bothalia* **14**, 745–758 (1983).
45. Sikes, R. S. & Gannon, W. L. & the Animal Care and Use Committee of the American Society of Mammalogists. Guidelines of the American Society of Mammalogists for the use of wild mammals in research. *J. Mammal.* **92**, 235–253 (2011).
46. Galen, M. et al. A multiplex panel of microsatellite markers for widespread sub-Saharan rodents of the genus *Mastomys*. *Mol. Ecol. Notes* **4**, 321–323 (2004).
47. Loiseau, A. et al. New polymorphic microsatellite loci for rodents of the genus *Mastomys* using PCR multiplexing, and cross-species amplification in *Myomys* and *Praomys*. *Mol. Ecol. Notes* **7**, 684–687 (2007).
48. Matschiner, M. & Salzburger, W. TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**, 1982–1983 (2009).
49. Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535–538 (2004).
50. Goudet, J. FSTAT (Version 1.2): a computer program to calculate F-statistics. *J. Hered.* **86**, 485–486 (1995).
51. Rousset, F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106 (2008).
52. Narum, S. R. Beyond Bonferroni: less conservative analyses for conservation genetics. *Conserv. Genet.* **7**, 783–787 (2006).
53. Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme F. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France, 1996–2004). Available at: [www.genetix.univ-montp2.fr](http://www.genetix.univ-montp2.fr).
54. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* **7**, 574–578 (2007).
55. Chiappero, M. A. et al. Contrasting genetic structure of urban and rural populations of the wild rodent *Calomys musculus* (Cricetidae, Sigmodontinae). *Mamm. Biol.* **76**, 41–50 (2011).
56. Dieringer, D. & Schlötterer, C. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* **3**, 67–169 (2003).
57. McRae, B. H. Isolation by resistance. *Evolution* **60**, 1551–1561 (2006).
58. Cushman, S. A., Wasserman, T. N., Landguth, E. L. & Shirk, A. J. Re-evaluating causal modelling with Mantel tests in landscape genetics. *Diversity*, **5**, 51–72 (2013).
59. Shirk, A. J., Wallin, D. O., Cushman, S. A., Rice, C. G. & Warheit, K. I. Inferring landscape effects on gene flow: a new model selection framework. *Mol. Ecol.* **19**, 3603–3619 (2010).
60. R Development Core Team R: *A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2010). Available at: <http://www.R-project.org>.
61. Wasserman, T. N., Cushman, S. A., Schwartz, M. K. & Wallin, D. O. Spatial scaling and multi-model inference in landscape genetics: *Martes americana* in northern Idaho. *Landsc. Ecol.* **25**, 1601–1612 (2010).
62. Oksanen, J. et al. vegan: community ecology package. R package version 2.3–4 (2015). Available at: <https://cran.r-project.org/web/packages/vegan/index.html>.
63. Bates, D. Maechler, M. & Bolker, B. lme4: linear mixed-effects models using 'Eigen' and S4. R package version 1.1–11 (2012). Available at: <https://cran.r-project.org/web/packages/lme4/index.html>.
64. Yang, R. C. A likelihood based approach to estimating and testing for isolation by distance. *Evolution* **58**, 1839–1845 (2004).
65. Clarke, R. T., Rothery, P. & Raybould A. F. Confidence limits for regression relationships between distance matrices: estimating gene flow with distance. *J. Agric. Biol. Environ. Stat.* **7**, 361–372 (2002).
66. Van Strien, M. J., Keller, D. & Holderegger, R. A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models. *Mol. Ecol.* **21**, 4010–4023 (2012).
67. Kutner, M., Nachsheim, C. & Neter, J. *Applied Linear Regression Models* (McGraw-Hill, New York, NY, 2004).
68. Rogerson, P. A. *Statistical methods for geography*. London, Sage (2001).
69. Barton, K. MuMIN: multi-model inference. R package version 1.15.6 (2016). Available at: <https://cran.r-project.org/web/packages/MuMIn/index.html>.
70. Johnson, J. B. & Omland, K. S. Model selection in ecology and evolution. *Trends Ecol. Evol.* **19**, 101–108 (2004).
71. Burnham, K. P. & Anderson, D. R. *Model Selection and Multi-Model Inference: A Practical Information Theoretic Approach* (New York, Springer-Verlag, 2002).
72. Oreluen, J. G. & Edwards, L. J. Fixed-effect variable selection in linear mixed models using statistics. *CSDA* **52**, 1896–1907 (2008).
73. Kenward, M. G. & Roger J. H. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983–997 (1997).
74. Halekoh, U. & Hojsgaard S. pbrtest: parametric bootstrap and Kenward Roger based methods for mixed model comparison. R package version 0.4–6 (2016). Available at: <https://cran.r-project.org/web/packages/pbrtest/index.html>.

## Acknowledgements

We are grateful to Ezemvelo KZN Wildlife for permitting us to do this work in the park (Dr. Ian Rushworth). We would also like to thank the staff at the Hluhluwe-iMfolozi Park for helping with the planning of the fieldwork, providing GIS maps and all other logistics (Dr. David Druce, Geoff Clinning and Cate James). Special thanks go to our field guide (Joseph Dlamini) for protecting us in the field and for always being willing to help. We are grateful to various family members and friends for assisting in the field. Stephen Spear is thanked for his constructive comments regarding analyses on earlier drafts of the manuscript. This study was funded by the National Research Foundation of South Africa (Postdoctoral Fellowship awarded to I.M.R. - SFP2008061100001) and Cardiff University.

### Author Contributions

I.M.R. and M.W.B. initiated the study and obtained research funding. All authors conceptualised the study. I.M.R. and C.L.S. collected specimens and I.M.R. was responsible for generating and interpreting the data. I.M.R. and M.B. performed data analyses. U.B. generated all GIS maps. All authors contributed to the writing and improvement of the manuscript. All authors read and approved the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Russo, I. M. *et al.* Landscape determinants of fine-scale genetic structure of a small rodent in a heterogeneous landscape (Hluhluwe-iMfolozi Park, South Africa). *Sci. Rep.* **6**, 29168; doi: 10.1038/srep29168 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>