# Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion

CÉCILE BERTHOULY-SALAZAR,*[1] CANG HUI,* TIM M. BLACKBURN,†‡ COLINE GABORIAUD,* BERNDT J. VAN RENSBURG,§[2] BETTINE JANSEN VAN VUUREN¶ and JOHANNES J. LE ROUX*

*Centre for Invasion Biology, Department of Botany & Zoology, Stellenbosch University, Private Bag X1, Stellenbosch 7602, South Africa, †Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK, ‡Distinguished Scientist Fellowship Program, King Saud University, Riyadh 1145, Saudi Arabia, §Department of Zoology and Entomology, Centre for Invasion Biology, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa, ¶Centre for Invasion Biology, Department of Zoology, University of Johannesburg, P.O. Box 524, Auckland Park 2006, South Africa*

## Abstract

**Conventional wisdom predicts that sequential founder events will cause genetic diversity to erode in species with expanding geographic ranges, limiting evolutionary potential at the range margin. Here, we show that invasive European starlings (*Sturnus vulgaris*) in South Africa preserve genetic diversity during range expansion, possibly as a result of frequent long-distance dispersal events. We further show that unfavourable environmental conditions trigger enhanced dispersal, as indicated by signatures of selection detected across the expanding range. This brings genetic variation to the expansion front, counterbalancing the cumulative effects of sequential founding events and optimizing standing genetic diversity and thus evolutionary potential at range margins during spread. Therefore, dispersal strategies should be highlighted as key determinants of the ecological and evolutionary performances of species in novel environments and in response to global environmental change.**

*Keywords*: genetic diversity, long-distance dispersal, range expansion, selection signature, invasion

## Introduction

Geographic range expansions are a natural consequence of population growth and dispersal, but the fate of genetic diversity during such expansions presents a conundrum for population and evolutionary biology (Austerlitz *et al.* 1997; Klopfstein *et al.* 2006; Travis *et al.* 2007; Excoffier *et al.* 2009; Petit 2011). Range expansions characterized by short-distance dispersal (SDD) will result in a reduction in genetic diversity in populations at the expanding range front, as these populations suffer from sequential founding events and genetic drift (Austerlitz *et al.* 1997). This reduction can be further enhanced by allele surfing, where new and existing mutations can reach high frequencies and thus high probabilities of fixation towards the range front (Klopfstein *et al.* 2006). Such allele surfing can further lead to migration loads through the fixation of deleterious mutations at the expanding range front (Travis *et al.* 2007; Excoffier *et al.* 2009), whilst beneficial mutations generally do not surf far from source regions (Travis *et al.* 2007). However, the persistence and surfing probabilities for all mutations will also depend on demographic processes (Münkemüller *et al.* 2011) and whether mutations directly affect dispersal (Travis *et al.* 2010). Therefore, erosion of genetic diversity due to successive founding events during range expansions may limit evolutionary potential at the range margin. In consequence, the expansion may be slowed by increased local-extinction risk and the disappearance of features of populations necessary to allow further spread.

Correspondence: Johannes J. Le Roux, Fax: +27 21 808 2995; E-mail: jleroux@sun.ac.za

[1]Present address: Institut de Recherche pour le Développement, UMR Diversité, Adaptation et Développement des Plantes (DIADE), Avenue Agropolis, BP 64501, 34394, Montpellier Cedex 5, France

[2]Present address: School of Biological Sciences, University of Queensland, Bribane, Qld 4072, Australia

The solution to the conundrum posed by the loss of evolutionary potential from sequential founding events at the expanding range margin may lie in particular dispersal strategies that natural selection will favour, and how populations at the range margin overcome sequential founding effects, especially as these populations constantly encounter novel environments. Specifically, the occurrence of long-distance dispersal (LDD) events could be important in mitigating the genetic constraints associated with spread via pure diffusion (i.e. SDD). While some have argued that gene flow and allele exchange via LDD from core populations could prevent local adaptation of peripheral populations and therefore limit the range expansion (Kirkpatrick & Barton 1997; Lenormand 2002), other models do show the potential, at least theoretically, for LDD to improve the fitness at the periphery even with only moderate gene flow (Kawecki 2000; Alleaume-Benharira *et al.* 2006; Kremer *et al.* 2012). Stratified dispersal (involving both SDD and LDD) has traditionally been modelled using thin-tailed leptokurtic dispersal kernels where LDD events are very rare. The outcome is similar to models exclusively involving SDD (i.e. pure diffusion): erosion of genetic diversity along the expansion path (Bialozyt *et al.* 2006). In contrast, fat-tailed dispersal kernels where LDD events are more frequent can in theory conserve population-level genetic diversity (Bialozyt *et al.* 2006) by allowing significant allele exchange between core and peripheral populations (Fayard *et al.* 2009), mitigating the effect of sequential founding events experienced by populations characterized by SDD (or a thin-tailed dispersal kernel) during range expansion. As genetic diversity and/or heterozygosity are proxies for adaptive potential and fitness (Guarino & Lobell 2011; Wetzel *et al.* 2012), frequent LDD events are expected to boost evolutionary potential in peripheral populations, which then can further accelerate the range expansion.

Examples of how LDD may have shaped population structure during postglacial colonization at large (continental) scales exist for plants (Petit *et al.* 1997, 2001, 2002; McLachlan *et al.* 2005). However, rapid and large-scale present-day expansions are rarely observed in native species, which are mostly in equilibrium with their environment, making it problematic to infer how LDD can affect contemporary evolutionary potential during range expansion. Nonindigenous populations introduced to novel environments are often characterized by small founding population sizes, with some experiencing rapid range expansion once established. Although environmental constraints on expansion are expected to differ between native and invasive species (Olivieri 2009), invasive alien species provide interesting model systems for studying how dispersal affects evolutionary potential. Here, we exploit just such a system by delineating the spatial genetic patterns of invasive, nonindigenous European starlings (*Sturnus vulgaris*) in South Africa. This population derived from a single introduction of <20 birds from England to Cape Town (Harrison *et al.* 1997) in the late 19th century, and since then has spread rapidly and largely unidirectionally eastwards from the area of release. The distribution of genetic variation in this starling population reveals how genetic diversity is conserved in an expanding population during the course of invasion and thus range expansion.

The first hypothesis tested here is that South African populations are genetically depauparate compared with native populations due to a single and small introduction event. To test this, we compare genetic diversities between native and invasive populations of European starlings to estimate the strength of the founder event in South Africa. Second, we test the hypothesis that genetic diversity is eroded along the expansion range of European starlings in South Africa. If dispersal is primarily diffusive, that is, SDD, then sequential founding events would reduce genetic diversity along the invasion pathway (Chakraborty & Nei 1977; Clegg *et al.* 2002). However, this genetic erosion could be mitigated if European starlings frequently disperse over long distances, that is, following a fat-tailed dispersal kernel. Lastly, we test whether abiotic factors influence, trigger or limit dispersal by relating genetic structure to environmental and habitat quality factors.

## Material and methods

### *Sampling design*

Models of landscape/spatial influence on genetic variation require continuously distributed sampling within the scale of spatial autocorrelation. We designed a random sampling protocol to capture the variability of landscape variables across the current distribution of European starlings in South Africa (available at sabap2.adu.org.za). Geographic information system (GIS) layers were used to identify the ranges of altitude and four bioclimatic variables (mean winter precipitation, mean summer precipitation, mean summer maximum temperature and mean winter minimum temperature) that are found across the current distribution of starlings in South Africa (SA), obtained from the WorldClim database (Hijmans *et al.* 2005). This allowed us to identify 50 sampling locations that captured the range of variation in these variables. Starlings were found at only 35 of these locations (Figs 1 and S1, Supporting Information), with a total of 232 individuals were sampled during the summer and autumn of 2011 (an average of 6.6

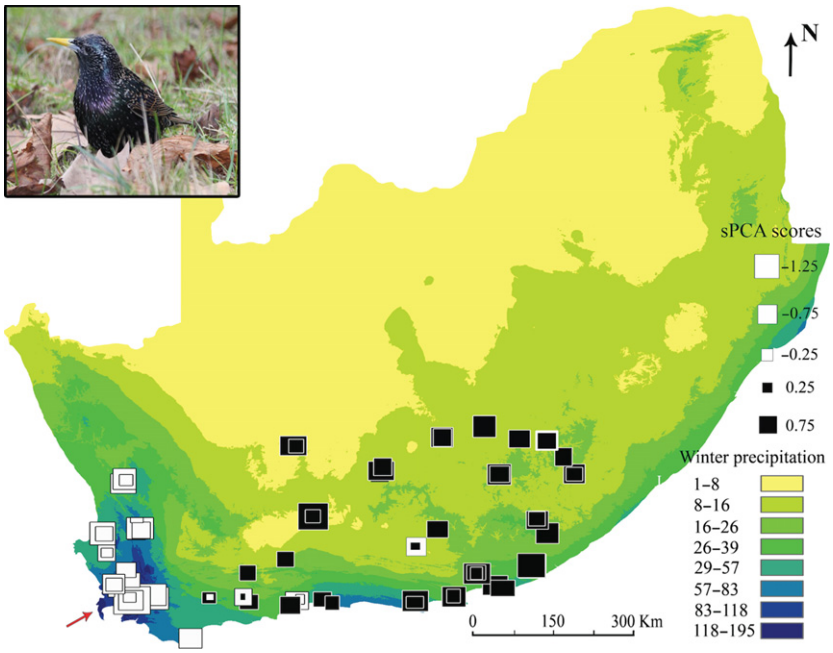



**Fig. 1** Spatial genetic pattern of European Starlings in South Africa. Gradient of winter precipitation (in mm) is shown. Red arrow indicates the site of introduction. Squares represent individual's scores for the 1st axis of the spatial PCA.

individuals per site, Table S1, Supporting Information). Note that we use 'population' and 'individuals at a site' interchangeably hereafter. For comparison, we also included samples from 16 individuals from the United Kingdom (UK). The UK specimens were all from the Garden Bird Health initiative (GBH*i*) archive. Specimens in this archive are dead birds found by members of the public and sent to the Institute of Zoology from across England (2/3) and Wales.

### *Molecular analyses*

Genomic DNA was extracted from tissue samples using the QIAGEN© Tissue extraction kit according to manufacturer's instructions (Qiagen). PCR products were labelled with fluorescent dyes and genotyped using a capillary sequencer (3730XL; Applied Biosystems) for 17 microsatellite markers previously designed for passerine species (Table S2, Supporting Information). Allele sizes were estimated using GENEMAPPER® version 3.7 (Applied Biosystems). For microsatellite data, the presence of null alleles was tested using FreeNA (Chapuis & Estoup 2007) and standard genetic indices estimated using GENEPOP (Raymond & Rousset 1995) and HP-RARE (Kalinowski 2005). For all subsequent analyses, we used two measures of genetic distance: $F_{ST}$ between pairwise populations (i.e. sampling sites) and Rousset's â for pairwise distances between individuals.

The mitochondrial DNA control region was sequenced for 1092 bp using primers svCRL1/svCRL2-svPhe3 (Rollins *et al.* 2011) and was aligned with 17 additional UK haplotypes from 27 individuals that originated from Monks Wood, Cambridgeshire, England (Rollins *et al.* 2011, GenBank Accession nos: H263626, H263630–42, FJ542128–9, FJ542133). We analysed the mitochondrial data using ARLEQUIN (Excoffier *et al.* 2005) and performed a minimum spanning tree from Kimura-2 parameter distances. Haplotypes from this study were deposited in GenBank (KF638591–617).

### *Spatial visualization of genetic diversity*

Spatial principal component analysis (sPCA; Jombart *et al.* 2008) was performed to investigate spatial patterns of genetic variability using the *adegenet* package in R (R Development Core Team 2010). sPCA scores can be informative at two major scales: global structures (i.e. large scales) display positive spatial autocorrelation between individuals, whereas local structures (i.e. fine scales) display negative spatial autocorrelation. The significance of global and local structures was tested using Monte Carlo tests as implemented in the *adegenet* package using 10 000 permutations. Our choice to investigate genetic spatial structure using sPCA stems from the fact that traditional Bayesian clustering approaches is an inappropriate method when individuals are genetically structured as a cline as in the case of isolation-by-distance (IBD; Guillot *et al.* 2009). Furthermore, sPCA allows the identification of IBD patterns (linearly decreasing scores) and genetic substructures (sudden changes in scores). Nevertheless, we also performed a Bayesian assignment analysis using STRUCTURE (Pritchard *et al.* 2000). For this analysis, we used the admixture model and correlated allele frequencies (Falush *et al.*

2003). Preliminary runs were performed from $K$ clusters = 1–10 with 300 000 iterations and a burn-in period of 100 000. Following preliminary results, five final runs for $K$ clusters = 1–5, with $10^6$ iterations following a burn-in period of 500 000 were performed.

### *Landscape connectivity*

We used an isolation-by-resistance approach (McRae 2006) to investigate which environmental factors might influence connectivity (i.e. gene flow). Landscape connectivity was evaluated using cost-distance modelling, whereby the geographic distance between locations is weighted by the values of the environmental factor tested, to estimate the likely difficulty of dispersal across that space expressed in 'environmental' (i.e. cost) terms. Each raster pixel of the landscape was assigned a resistance for a specific environmental feature (i.e. topographic or bioclimatic factors), representing the difficulty for dispersing due to this environmental feature. From this resistance (or cost) surface, the effective cost distance between two locations was calculated using the least-cost path estimated by PATHMATRIX software (Ray 2005). Correlations between genetic and cost distances were tested to identify which environmental features affect gene flow (i.e. dispersal) across the landscape.

Specifically, for the five environmental factors, we first examined the distribution histograms of each factor across the South African distribution range of starlings. From these, we created 6–10 classes to which we attributed resistance values ranging from 0.1 for the minimum resistance to 1 for the maximum resistance. The most common environmental class was assigned the smallest resistance value and the rarest class the highest resistance. We then searched for the least-cost path.

We also verified that Rousset's â distances followed a Gaussian distribution, and geographic and cost distances were ln-transformed to reduce their skewness and to conform to the assumptions of Mantel tests. We tested the correlation between individual Rousset's â and cost distances after correction for geographic distance using partial Mantel tests. Analyses were performed with 10 000 permutations using the ZT software (Bonnet & Van de Peer 2002).

### *Changes of genetic parameters along the invasion pathway*

We first tested whether a correlation exists between individual Rousset's â and geographic distances using Mantel tests, that is, the IBD pattern across the entire range (IBD slope $b_{SA}$). However, changes in IBD patterns may occur during range expansion and may reflect demographic processes or changes in dispersal strategies. To this end, we also investigated changes in IBD along the invasion pathway using a slide-window analysis (Castric & Bernatchez 2003). A constant-width window was slid along the direction from Cape Town to the most eastward population found (i.e. invasion pathway), which lies 1053 km from Cape Town. Therefore, in order to include a minimum of four populations in each window and to maximize the number of windows, a window width of 200 km and a slide distance of 20 km were chosen. Still, not every slide step included a new site, and any that did not were therefore removed from the analysis, leaving a total of 17 windows. Within each window, a least-squares regression between population's genetic distances [$F_{ST}/(1-F_{ST})$] and log-transformed geographic distances was performed and the regression slope estimated. To be consistent across windows, a rarefaction approach was applied by resampling four sites for all windows containing more than four sites. Change in IBD was analysed by plotting the slope values for the 17 windows along the invasion pathway ($b_{Col}$). To investigate whether the observed relationship between regression slope and distance could be obtained by chance, we performed new analyses by randomly reshuffling sampling sites 1000 times along the invasion pathway. Parameters of the simulated relationships between IBD slopes and linear distances along coastlines were compared against the observed parameters for significance. For allelic richness (AR), expected ($H_E$) and observed ($H_O$) heterozygosity, we estimated the window means using the rarefaction approach of four sites per window and regressed them with the distance of the window from Cape Town.

### *Environmental effects*

Spatial Moran's eigenvector map (MEM) variables, derived from geographic coordinates, were used to investigate distribution patterns of allele frequencies in response to environmental factors (Manel *et al.* 2012). Moran's eigenvector map (MEM) variables have been shown to be powerful descriptors of spatial patterns and reveal spatial and environmental effects on community assembly, genetic variation (Manel *et al.* 2012) or morphometric variation (Berthouly-Salazar *et al.* 2012). Briefly, we regressed the frequencies of each of the 119 identified alleles in the South African starling population against five environmental factors and 15 spatial MEMs for considering spatial autocorrelation. Alleles for which multivariate regression was significant and showed an adjusted $R^2 > 0.5$ were considered as potentially ecologically relevant. For these

alleles, variables that had a significant effect were tested using univariate regressions. Calculations were carried out using the *PCNM* package in the R statistical environment (R Development Core Team 2010).

## Results

### *Genetic diversity*

We sequenced 1092 bp of the mitochondrial control region and identified 37 different haplotypes. For the global data set (UK+SA), mean pairwise differences between haplotypes were on average 5.2 ± 2.6 (SD) while nucleotide diversity averaged 0.006 ± 0003. Across all haplotypes, we identified 35 mutations of which 19 were parsimony informative sites and 16 singleton variable sites. Due to this low informativeness, we were not able to reconstruct a phylogeny. From our 16 UK samples, we identified 13 unique haplotypes which, in combination with the data of Rollins *et al.* (2011) led to a total of 26 UK haplotypes. We found 17 haplotypes in South Africa among which six were shared with the 26 UK haplotypes. Three of the six shared haplotypes between South Africa and the UK were present in 70% of the South African sampling localities. Using mitochondrial DNA, estimates of $\theta_k$ suggest that effective population size of native range starlings in the United Kingdom was 6.5 times larger than in South Africa. Four main haplotypes in South Africa represent 88% of the population. With the exception of four singletons from the western part and four singletons from the eastern part of the distribution, all other haplotypes were distributed across the entire invasive range (Fig. 2). A map of distribution of each haplotype shows that some haplotypes such as Hap45 were only found at the core and at the periphery of the invasive range, suggesting long-distance dispersal (Fig. S2, Supporting Information).

No microsatellite markers were found with potential null alleles. The average number of alleles per marker was 7.5, ranging from two for Patmp2-43 to 15 for Sta213 and Sta308 (Table S2, Supporting Information). Only two alleles were found out of Hardy–Weinberg equilibrium after Bonferroni correction (Ase19 and SS2-106).

At the country level, we found that expected heterozygosity ($H_E$) and rarefied allelic richness (AR) in South Africa to be similar to those found in the United Kingdom ($H_E$: 0.653 vs. 0.698; AR: 5.4 vs. 6.2). Within South Africa, $H_E$ ranged from 0.539 to 0.682 and $H_O$ from 0.494–0.718 (Table S1, Supporting Information). Overall, we found a low inbreeding rate throughout South Africa ($F_{IS}$ = 0.077).

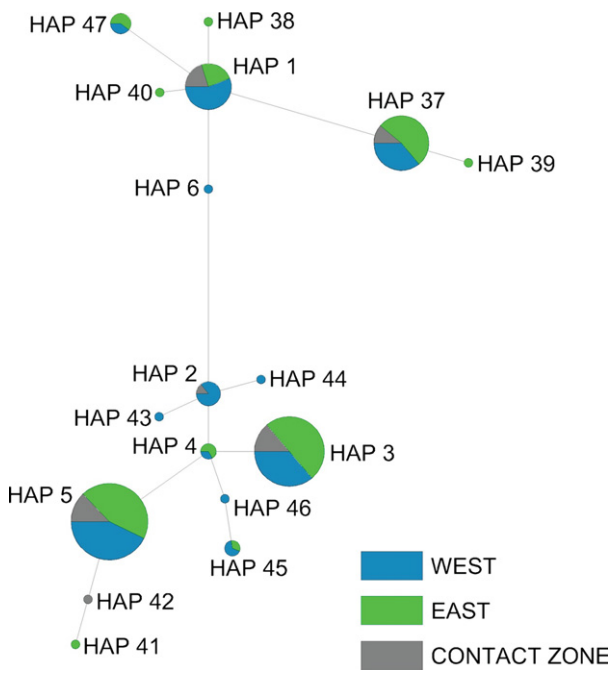



**Fig. 2** Minimum spanning tree of 17 South African haplotypes using K2P distances. Circles are proportional to the frequency of each haplotype. Haplotypes labelled as 'contact zone' include the following sampling sites: EWC6, EWC7 and EWC8 (see Fig. S1, Supporting Information).

### *Spatial patterns*

For South African starlings, the correlation between genetic distances (Rousset's â) and geographic distances (LnG) showed only a weak IBD pattern ($b_{SA}$ = 0.028, $P$ = 0.054). However, our sliding window approach indicated a decline in IBD slopes along the eastward progression of the range [($F_{ST}$/(1−$F_{ST}$) vs. LnG; $b_{Col} = -3 \times 10^{-5}$, $R^2$ = 0.55, outside the 95% CI ($-8.3 \times 10^{-6}$; $9.85 \times 10^{-6}$), Figs 3A and S3, Supporting Information]. In addition, this approach also revealed a constant $H_E$ across the entire invasive range ($b_{HE} = 3 \times 10^{-6}$, $R^2$ = 0.1, $P$ = 0.65) but a decrease in rarefied AR ($b_{AR} = -1 \times 10^{-04}$, $R^2$ = 0.32, $P < 0.05$; Fig. 3B) and an increase in $H_O$ ($b_{Ho} = 2 \times 10^{-05}$, $R^2$ = 0.68, $P < 0.001$ Fig. 3C).

When we compared the frequencies of alleles lost between the first and the last slide window, we found that the mean frequency of the 20 lost alleles (0.05) to be significantly lower than expected from the permutation test (10 000 reshuffles; mean = 0.156, $P < 0.0001$). In contrast, for those alleles kept, the mean frequency (0.18) was significantly higher than expected from the permutation test (mean = 0.156, $P < 0.0001$). We also found lower levels of genetic differentiation between peripheral populations compared with core populations ($F_{ST}$ = 0.0318 for populations 0–150 km to the introduc-

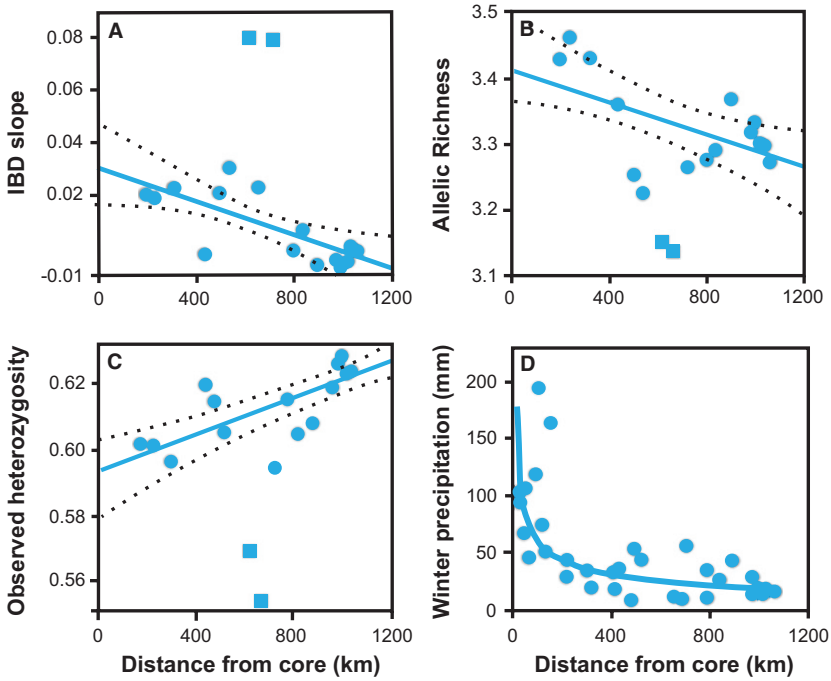



**Fig. 3** Pattern changes along the invasion pathway. Relationships between geographic distance (km) from core (i.e. the introduction site in Cape Town) A) IBD slopes, B) allelic richness (AR), C) observed heterozygosity ($H_O$), as determined by slide-window analysis and D) winter precipitation (mm) of sampling sites. Regression lines are indicated in blue with 95% CI in dotted black lines. Closed squares were considered as outliers (1 population very different from others in the same window) and were not taken into account to calculate *P*-values in regressions analyses.

tion site; $F_{ST} = 0.0127$ for populations 800–1050 km away; Mann–Whitney test, $P < 0.001$).

For the sPCA analysis, global structures (i.e. large spatial scale) were significant ($P = 0.0073$) with two first strong eigenvalues (Fig. S4, Supporting Information) while local structures were not significant ($P = 0.243$). The first axis of the sPCA identified two subpopulations with a contact zone around 150–300 km east from the introduction site (Fig. 1). Similar breakdown was found for the Bayesian assignment tests conducted in STRUCTURE (Appendix S1, Supporting Information). However, as genetic differentiation is very low between these two subpopulations ($F_{ST} = 0.008$), this genetic substructure most likely reflects recent divergence. Both subpopulations have low IBD slopes, although the population to the west of the contact zone has a steeper pattern of IBD ($b_W = 0.104$, $P < 0.001$) compared with the eastward population ($b_E = 0.054$, $P < 0.001$).

### *Environmental effects*

Sixteen of the 119 multivariate regressions of allele frequencies against MEMs and environmental factors showed signs of ecological relevance, as assessed by their significance ($P < 0.05$) and variance explained ($R^2 > 0.5$). Following univariate regressions for each of these 16 alleles, we found that three alleles restricted to few sampling sites (SS2-71B_326, Sta213_173 and SS3-42C_135) and one widespread allele (Sta308_140, occurring in 25 of the 35 sites) were correlated with winter precipitation only (Table 1, Fig. 3D).

Our isolation-by-resistance approach indicated only two significant correlations between genetic distances and environmental cost distances (Table 2). Genetic distances (â) increased with higher winter precipitation (partial Mantel test: $b_{Wp} = 0.051$, $P < 0.001$) and decreased with higher summer precipitation ($b_{Sp} = -0.056$, $P < 0.001$), and thus, gene flow is limited where winter precipitation is high and summer precipitation is low.

## Discussion

Biological invasions are typically characterized by stochasticity, severe founder events, and strong genetic drift, so that the genetic diversity found in introduced populations usually only represent a small proportion of that found in the native range (but see, Dlugosch & Parker 2008). As adaptive potential is linked to genetic diversity (Guarino & Lobell 2011), low genetic diversity

**Table 1** Regression values for all significant univariate regressions per locus

| Locus_allele | Environmental factor | Slope | SE | *t*-value | *P*-value | $R^2$ adjusted |
|---|---|---|---|---|---|---|
| SS2-71B_326 | Winter precipitation | 0.0005805 | 0.0001473 | 3.940 | 0.000398 | 0.30 |
| Sta213_173 | Winter precipitation | 0.0005705 | 0.0001473 | 3.872 | 0.000483 | 0.29 |
| SS3-42C_135 | Winter precipitation | 0.0006786 | 0.0001257 | 5.400 | $5.67e^{-06}$ | 0.45 |
| Sta308_140 | Winter precipitation | 0.0011726 | 0.0003259 | 3.598 | 0.00104 | 0.26 |

**Table 2** Results for partial Mantel tests between Rousset's â and environmental cost distances

| Environmental factor | Correlation value | *P*-value |
|---|---|---|
| Altitude | −0.044 | 0.0954 |
| Winter precipitation | 0.051 | 0.0005 |
| Summer precipitation | −0.056 | 0.0001 |
| Tmax in summer | 0.0079 | 0.1135 |
| Tmin in winter | −0.036 | 0.0937 |

Tmax: maximum temperature; Tmin: minimum temperature.

in introduced populations may reduce their abilities to adapt and hence increase their extinction risk in novel environments (Hoffmann & Blows 1994).

The colonization history of European starlings in South Africa has been well documented, with the successful establishment of a founder population in Cape Town around 1930 following the introduction of <20 individuals in late 1890s (Harrison *et al.* 1997 and references therein, Picker & Griffiths 2011). This was followed by a clear eastward range expansion that reached the Eastern Cape Province (i.e. East-London) by 1960, and finally the Kwazulu Natal Province (i.e. Durban) by the beginning of 2000 (Fig. S1, Supporting Information Hui *et al.* 2012). Our results showed that mitochondrial haplotype diversity of this invasion is surprisingly high given a founding population of <20 birds (Harrison *et al.* 1997). The seventeen haplotypes identified in South Africa suggest either a heavily sex-biased introduction (towards females) occurred or that new mutations happened *in situ* following introduction, both of which seem highly unlikely. Alternative explanations could be that an undocumented secondary introduction(s) followed the initial introduction to Cape Town (Long 1981) or that the original introduction to Cape Town must have been larger than 20 birds. The eastward expansion of European starlings from Cape Town never showed a discontinuous distribution (Hui *et al.* 2012; Fig. S1, Supporting Information), and therefore, a secondary introduction(s), if any, could only have occurred in areas already occupied by the expanding founding population. Moreover, the probability of newly arriving haplotypes to persist in large already established populations would be low due to the effects of drift, and therefore, these secondary introductions must have occurred early on in the invasion or at peripheral low-density populations. Nevertheless, despite the unexpectedly high number of haplotypes, our data do suggest a strong founding effect as shown by the $\theta_k$ ratio. Estimating the loss of genetic diversity in introduced populations can be a daunting task, requiring sufficient sampling in both native and invasive ranges. According to Puillandre *et al.*'s (2008), the sample sizes for both ranges used here are sufficient to infer that the South African population has been subjected to a founder effect.

Unlike for our mitochondrial data, the founding event of European starlings in South Africa had little effect on microsatellite allelic richness with very low differences observed between the native UK and South African invasive populations. Evidence of equal or higher genetic diversity in species' invasive ranges has been explained by admixture following multiple introductions, especially when introductions occur from different native range regions. Dlugosch & Parker (2008) not only observed that allelic richness, and not necessarily heterozygosity, would benefit from multiple introductions, but also that multiple introductions may not always lead to higher genetic diversity in invasive populations. In addition, many extrinsic and intrinsic factors during the invasion process will impact genetic diversity, irrespective of the amount of genetic diversity introduced. In other words, multiple introductions do not always rescue invaders from loss of diversity and others factors may affect levels of diversity following species introductions, one being dispersal strategies.

Dispersal strategies will have important implications for the fate of genetic diversity during the rapid range expansions typically associated with successful invasive species, yet their effects have not been explicitly studied. Dispersal by pure diffusion (SDD only) can increase genetic differentiation along the expansion range (Hallatschek *et al.* 2007) while LDD will potentially attenuate genetic differentiation and therefore the IBD patterns (Bialozyt *et al.* 2006; Fayard *et al.* 2009; Ray & Excoffier 2010). At both global (i.e. entire invasive range) and local (i.e. subpopulations) scales, South

African starling populations showed very weak IBD, which may simply be the effect of short residence times. However, assuming a generation time of 1.6–2 years (Feare & Forrester 2002), these populations arrived between 56 and 70 generations ago, which can be enough time for IBD to develop. In fact, since their introduction, two slightly differentiated starling subpopulations have already emerged (Fig. 1). In addition, our results also showed an absence of spatial structure in mtDNA, with core and peripheral populations sharing a significant number of haplotypes (Fig. 2), including low frequency haplotypes (e.g. Hap45; Fig. S2, Supporting Information). Therefore, our results seems to indicate that the weak IBD pattern observed means that LDD events are relatively frequent in the invasive starling population in South Africa.

Bialozyt *et al.* (2006) simulated the effects of LDD on maternally inherited haploid genetic diversity at the wave front and found that diversity increased when LDD events reached a propability of $\geq 10^{-5}$. On the other hand at probabilities of $<10^{-5}$, diversity levels decreased below those expected under pure SDD models, a phenomenon known as the embolism effect (Petit *et al.* 2004). This nonmonotonous response can be further amplified in large populations. Increased frequencies of LDD events (i.e. fat-tailed dispersal kernels) would preserve genetic diversity by allowing gene flow across the whole range, the so-called 'reshuffling effect' (Ray & Excoffier 2010). Therefore, frequent LDD can increase diversity along the expansion range (Fayard *et al.* 2009), while simultaneously limiting allele surfing and therefore the effect of drift on genetic differentiation.

Studies on how the frequency of LDD affects genetic diversity in natural populations are rare. Recently, Szövényi *et al.* (2012) studied different species of long-distance-dispersing peat mosses. They found no or very weak IBD patterns and mainly positive, but not significant, correlations between allelic richness or $H_E$ and distance from the core. If the theoretical predictions discussed above are correct, then fat-tailed dispersal kernels would allow species to conserve most genetic diversity, or may even increase genetic diversity along the expansion range. Here, we observed a weak but significant decrease in AR along the expansion range. This was partly explained by the loss of rare alleles and is not in contradiction with the effect of frequent LDD as LDD do not entirely prevent the cumulative effect of founding events (Austerlitz *et al.* 1997; Comps *et al.* 2001). Allelic richness is expected to be lost more easily from bottlenecks than heterozygosity (Nei *et al.* 1975). Moreover, LDD limits allele surfing, a process by which rare alleles can be conserved and even reach high frequencies at the expansion front as described by the embolism effect.

We found a positive correlation between heterozygosity and distance from the core (but only significant for $H_O$) and lower genetic differentiation among peripheral populations than among core populations. These observations would be in agreement with the 'reshuffling effect' expected with frequent LDD events. LDD events mix individuals from different sources, thus increasing genetic diversity while decreasing genetic differentiation. Increasing genetic diversity at the expanding front may represent the advantage of maximizing adaptive potential where populations face new environmental conditions.

Dispersal is often viewed as a static life history trait. This is probably unrealistic as indicated by Reid's paradox (Clark *et al.* 1998), which notes that species ranges often expand much faster than expected from normally observed dispersal rates. This paradox was solved by theory in showing that the rate of dispersal of individual animals and plants should increase towards the front of an expanding geographic range (Le Corre *et al.* 1997; Travis & Dytham 2002). Subsequently, Phillips *et al.* (2006) were the first to show an increase in dispersal abilities towards the expanding range front for invasive cane toads. Since then, others have confirmed that accelerating dispersal can evolve during the course of range expansion through 'spatial sorting' of individuals with greater dispersal abilities (Berthouly-Salazar *et al.* 2012) and that individuals' dispersal strategies depend on a series of internal and external conditions (Clobert *et al.* 2009). A recent study also showed that kin competition (i.e. relatedness) can change the shape of the dispersal kernel towards more 'fat-tailed' kernels (Bitune *et al.* 2013).

Historical occurrence records indicate that European starlings have accelerated their dispersal during their invasion in South Africa (Hui *et al.* 2012). Our genetic results showing a decrease in IBD slopes along the axis of range expansion corroborate these inferences of enhanced dispersal towards the periphery (Fig. 3A). Lower IBD slopes observed for windows at the expanding front could be due to not enough time having lapsed since arrival for an IBD pattern to emerge. However, as LDD events seem evidently to occur and given the observed lower IBD slope in the eastward population compared with the westward population, we believe that an acceleration of dispersal occurred. This is in agreement with historic occurrence records indicating a two-fold increase in the speed of range expansion 150 km east of Cape Town (Hui *et al.* 2012), the approximate contact zone between the two subpopulations we identified. Interestingly, this also represents an area where climate conditions, especially winter precipitation, change dramatically (Fig. 3D). Environmental quality is likely to influence dispersal

(Dytham 2009; Hui *et al.* 2012) due to costs (if the recipient environment is less favourable than the original) or benefits (if it allows individuals to escape from poor environments). Using two types of analyses, we found that starling dispersal is indeed affected by environmental factors, especially winter precipitation. First, we found signatures of selection at four microsatellite alleles. While three alleles were only present in a few sites and therefore potentially represent false-positive effects due to stochasticity of founding events, one allele was found across the invasive distribution range. Detecting signatures of selection across a species range can be very difficult when allelic clines resulting from IBD correlate with spatial and environmental scales (Holderegger *et al.* 2010). However, we did not observe a significant IBD pattern across the whole range and only very weak IBD patterns at the subpopulation level. Moreover, the approach used here has considered spatial autocorrelation via MEM variables, thus the probability of detecting false selection signatures is unlikely. Second, our isolation-by-resistance approach shows that gene flow is influenced by precipitation. We found an increase in genetic distances, and thus a decrease in dispersal, as winter precipitation increases. Similarly, Hui *et al.* (2012) estimated much lower spread rates (4–8 km/year) for European starlings in areas with high winter precipitation compared with eastward regions in South Africa where winter precipitation is lower (8–32 km/year). Precipitation is likely to be a key factor defining the environmental niche of starlings in the region and often is a surrogate for other environmental factors, in particular for primary production and therefore resource availability. In fact, it has been previously found that precipitation is a key factor in determining species richness for many native South African birds (van Rensburg *et al.* 2002).

Overall, therefore, our study shows that as habitat becomes less suitable to the east, the more challenging rainfall regime encountered by starlings triggers the acceleration of dispersal. The optimum habitat of starlings in South Africa is situated within a 150–200 km radius from Cape Town (the introduction site). Favourable habitats along the southeast coast, coupled with relatively high human densities, have formed a corridor that has directed a primarily eastward pathway of range expansion. Subsequent contact with less favourable habitat by the range front (which occurred around 1940) triggered faster expansion (Hui *et al.* 2012), as reflected by the lower IBD slope found in the eastward population, supporting the 'good-stay, bad-disperse' rule in response to environmental conditions previously identified for starlings in Britain (Hui *et al.* 2012).

## Conclusion

The spread of European starlings in South Africa is characterized by relatively frequent LDD events, which mitigate the effects of sequential founding events during range expansion and allow the conservation of genetic diversity. LDD events redistribute genetic diversity across the expanding range and can transport beneficial mutations to peripheral populations from anywhere across the species' range (Fayard *et al.* 2009), increasing evolutionary potential, as shown previously for other species that are shifting their native ranges due to climatic changes (Buckley *et al.* 2012). LDD would be selected for in populations facing intermediate local-extinction pressure (Bohrer *et al.* 2005). Here, we suggest that dispersal strategy, and specifically enhanced dispersal, can therefore be regarded as an evolutionary response to avoid the loss of genetic diversity during rapid range expansions. Our analyses highlight how unfavourable environmental conditions can lead to enhanced dispersal that further allows for the conservation of genetic diversity by optimizing genetic diversity and thus evolutionary potential.

In conclusion, (i) the European starling follows a 'good-stay, bad-disperse' strategy with its optimum niche sensitive to variation in rainfall; (ii) starlings encountering unfavourable habitat at the range front around 1940 were a trigger for more frequent LDD events and faster expansion as a result; and (iii) a relatively high proportion of LDD events help to preserve genetic diversity, especially at the range front to enhance fitness in novel environments. This may not only be an important mechanism for successful invasions by nonindigenous species, but also shows in general how dispersal strategy may help species cope with climate or other environmental change, especially if that change leads to a deterioration of habitat conditions in a previously favourable location.

## Acknowledgements

We thank Danie du Preez, Ethan Newman and Jaco Visser and many others for their assistance in sampling specimens, Lee Ann Rollins for providing her data on European Starlings and colleagues at Stellenbosch University for valuable inputs on earlier drafts of the manuscript. We thank the members of the public and participants in the British Trust for Ornithology's Garden BirdWatch who submitted carcasses to the GBH*i*, and Becki Lawson for help obtaining the relevant samples. We thank three anonymous reviewers from their useful comments. CB-S received a postdoctoral grant from the South African National Research Foundation (NRF). This study was made possible through the financial support from Stellenbosch University Subcommittee B (JLR and CH) and the National Geographic Society Grant N°8492-11 (CB-S). The animal care committee of Stellenbosch University (N°10NP_CEC01) and CapeNature (N°AAA-004-00499-0035) approved the protocol for this study.

## References


Alleaume-Benharira M, Pen IR, Ronce O (2006) Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *Journal of Evolutionary Biology*, **19**, 203–215.

Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H (1997) Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, **51**, 148–164.

Berthouly-Salazar C, van Rensburg B, Le Roux J, van Vuuren BJ, Hui C (2012) Spatial sorting drives morphological variation in the invasive bird, *Acridotheris tristis*. *PLoS ONE*, **7**, e38145.

Bialozyt R, Ziegenhagen B, Petit RJ (2006) Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology*, **19**, 12–20.

Bitune EV, Bonte D, Ronce O, *et al.* (2013) Density and genetic relatedness increase dispersal distance in a subsocial organism. *Ecology Letters*, **16**, 430–437.

Bohrer G, Nathan R, Volis S (2005) Effects of long-distance dispersal for metapopulation survival and genetic structure at ecological time and spatial scales. *Journal of Ecology*, **93**, 1029–1040.

Bonnet E, Van de Peer Y (2002) zt: a software tool for simple and partial Mantel tests. *Journal of Statistical Software*, **7**, 1–12.

Buckley J, Butlin RK, Bridle JR (2012) Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Molecular Ecology*, **21**, 267–280.

Castric V, Bernatchez L (2003) The rise and fall of isolation by distance in the anadromous brook charr (*Salvelinus fontinalis* Mitchill). *Genetics*, **163**, 983–996.

Chakraborty R, Nei M (1977) Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution*, **31**, 347–356.

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

Clark JS, Fastie C, Hurtt G *et al.* (1998) Reid's paradox of rapid plant migration. *BioScience*, **48**, 13–24.

Clegg SM, Degnan SM, Kikkawa J, Mortiz C, Estoup A, Owen IPF *et al.* (2002) Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 8127–8132.

Clobert J, Le Gaillard J-F, Cote J, Meylan S, Massot M *et al.* (2009) Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecology Letters*, **12**, 197–209.

Comps B, Gömöry D, Letouzey J, Thiébaut B, Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the Eurpean beech. *Genetics*, **157**, 389–397.

Dlugosch KM, Parker I (2008) Founding events in species invasions: genetic variation, adaptive evolution and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.

Dytham C (2009) Evolved dispersal strategies at range margins. *Proceedings Biological sciences/The Royal Society*, **276**, 1407–1413.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver 30: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics*, **40**, 481–501.

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

Fayard J, Klein EK, Lefevre F (2009) Long distance dispersal and the fate of a gene from the colonization front. *Journal of Evolutionary Biology*, **22**, 2171–2182.

Feare CJ, Forrester GJ (2002) The dynamics of a suburban nestbox breeding colony of starlings *Sturnus vulgaris*. In: *Investigation Into the Causes of the Decline of Starlings and House Sparrows in Great Britain* (eds Crick HPQ, Robinson RA, Appleton GF, Clark NA, Rickard AD), pp. 73–90. DEFRA, Bristol.

Guarino L, Lobell DB (2011) A walk on the wild side. *Nature Climate Change*, **1**, 374–375.

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology*, **18**, 4734–4756.

Hallatschek O, Hersen P, Ralanathan S, Nelson DR (2007) Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 19926–19930.

Harrison JA, Allan DG, Underhill LG *et al.* (1997) *The Atlas of Southern African Birds*. BirdLife South Africa, Johannesburg.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hoffmann AA, Blows MW (1994) Species borders: ecological and evolutionary perspectives. *Trends in Ecology & Evolution*, **9**, 223–227.

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends in Plant Science*, **15**, 675–685.

Hui C, Roura-Pascual N, Brotons L, Robinson RA, Evans KL (2012) Flexible dispersal strategies in native and non-native ranges: environmental quality and the "good–stay, bad–disperse" rule. *Ecography*, **35**, 1024–1032.

Jombart T, Devillard S, Dufour A-B, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

Kalinowski ST (2005) hp-rare 10: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kawecki TJ (2000) Adaptation to marginal habitats: contrasting influence of the dispersal rate on the fate of alleles with small and large effects. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **267**, 1315–1320.

Kirkpatrick M, Barton NH (1997) Evolution of a species's range. *The American Naturalist*, **150**, 1–23.

Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, **23**, 482–490.

Kremer A, Ronce O, Robledo-Arnuncio JJ *et al.* (2012) Long-distance gen flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, **15**, 378–392.

Le Corre V, Machon N, Petit RJ, Kremer A (1997) Colonization with long-distance seed dispersal and genetic structure of

maternally inherited genes in forest trees: a simulation study. *Genetical Research*, **69**, 117–125.
Lenormand T (2002) Gene flow and the limits of natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.
Long JL (1981). Introduced birds of the world: the worldwide history, distribution and influence of birds introduced to new environments. *Universe Book*. David & Charles Publishers, Australia, 528 pp.
Manel S, Gugerli F, Thuillier W *et al.* (2012) Broad scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology*, **21**, 3729–3738.
McLachlan JS, Clark JS, Manos PS (2005) Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, **86**, 2088–2098.
McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.
Münkemüller T, Travis MJ, Burton OJ, Schiffers K, Johst K, *et al.* (2011) Density regulated population dynamics and conditional dispersal alter the fate of mutations occurring at the front of an expanding population. *Heredity*, **106**, 678–689.
Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.
Olivieri I (2009) Alternative mechanism of range expansion are associated with different changes of evolutionary potential. *Trends in Ecology & Evolution*, **24**, 289–292.
Petit RJ (2011) Early insights into the genetic consequences of range expansions. *Heredity*, **106**, 203–204.
Petit RJ, Pinneau E, Demesure B, Bacillieri R, Ducousso A, Kremer A *et al.* (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9996–10001.
Petit RJ, Bialozyt R, Brewer S, Cheddadi R, Compte B (2001) From spatial patterns of genetic diversity to postglacial migration processes in forest trees. In: *Integrating Ecology and Evolution in a Spatial Context*, BES symposium volume (eds Silvertown J, Antonovics J), pp. 295–318, Blackwell, Oxford.
Petit RJ, Csaiki UM, Bordacs S (2002) Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156**, 5–26.
Petit RJ, Bialozyt R, Garnier-Gere P, Hampe A (2004) Ecolgoy and genetics of tree invasions: from recent introductions to Quaternary migrations. *Forest Ecology and Management*, **197**, 117–137.
Phillips BL, Brown GP, Webb JK, Shine R (2006) Invasion and the evolution of speed toads. *Nature*, **439**, 803.
Picker MD, Griffiths CL (2011) *Alien and Invasive Animals – A South African Perspective*. Struik-Random House Publishers, Cape Town, 248 pp.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
Puillandre N, Dupas S, Dangles O *et al.* (2008) Genetic bottleneck in invasive species: the potato tuber moth adds to the list. *Biological Invasions*, **10**, 319–333.
R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available from http://wwwR-projectorg/.
Ray N (2005) Pathmatrix: a geographical information system tool to compute effective distances among samples. *Molecular Ecology Notes*, **5**, 177–180.
Ray N, Excoffier L (2010) A first step towards inferring levels of long-distance dispersal during past expansions. *Molecular Ecology*, **10**, 902–914.
Raymond M, Rousset F (1995) GENEPOP (version 12): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
van Rensburg BJ, Chown SL, Gaston KJ (2002) Species richness, environmental correlates, and spatial scale: a test using South African birds. *The American Naturalist*, **5**, 566–577.
Rollins LA, Wolnough AP, Sinclair R, Mooney NJ, Sherwin WB (2011) Mitochondrial DNA offers unique insights into invasion history of the common starling. *Molecular Ecology*, **20**, 2307–2317.
Szövényi P, Sundberg S, Shaw AJ (2012) Long-distance dispersal and genetic structure of natural populations: an assessment of the inverse isolation hypothesis in peat mosses. *Molecular Ecology*, **21**, 5461–5472.
Travis JMJ, Dytham C (2002) Dispersal evolution during invasions. *Evolutionary Ecology Research*, **4**, 1119–1129.
Travis JMJ, Mönkemöller T, Burton JOJ, Best A, Johst K (2007) Deleterious mutations can surf to high densities on the wave front of an expanding population. *Molecular Biology and Evolution*, **24**, 2334–2343.
Travis JMJ, Münkemüller T, Burton OJ (2010) Mutation surfing and the evolution of dispersal during range expansion. *Journal of Evolutionary Biology*, **23**, 2656–2667.
Wetzel DP, Stewart IR, Westneat DF (2012) Heterozygosity predicts clutch and egg size but not plasticity in a house sparrow population with no evidence of inbreeding. *Molecular Ecology*, **21**, 406–420.

---

C.B.S. participate in the design, sampling, analysis and writting. C.H. participate in the design, analysis and writting. T.B. participate in the sampling and writting. C.G. participate in the analysis. B.V.R. and B.V.V. participate in the design and writting. J.L.R. participate in the design, analysis and writting.

---

## Data accessibility

Mitochondrial haplotypes have been submitted to GenBank (KF638591–617).

A file containing all information on individuals genotyped, including location and genotypes at each locus, has been submitted to Dryad doi:10.5061/dryad.3bv1h.

Supplementary files:

Table S1. Summary of sampling sites and population genetic indices for each site. n: number of individuals, $H_E$: unbiaised expected heterozygosity; $H_O$: observed heterozyosity; A:mean number of alleles; Ar: allelic richness rarefied to n=4.

| Sampling site | Province | n | $H_E$ | $H_O$ | A | Ar |
|---|---|---|---|---|---|---|
| CT-1 | Western cape | 4 | 0.656 | 0.618 | 3.4 | 3.4 |
| CT-2 | Western cape | 10 | 0.645 | 0.682 | 4.5 | 3.4 |
| CT-3 | Western cape | 9 | 0.638 | 0.552 | 4.2 | 3.3 |
| CT-6 | Western cape | 21 | 0.637 | 0.565 | 5.5 | 3.5 |
| EWC-2 | Western cape | 6 | 0.607 | 0.588 | 3.9 | 3.4 |
| EWC-4 | Western cape | 11 | 0.655 | 0.589 | 4.8 | 3.5 |
| EWC-5 | Western cape | 6 | 0.595 | 0.591 | 4.2 | 3.5 |
| EWC-6 | Western cape | 18 | 0.641 | 0.611 | 5.5 | 3.4 |
| EWC-7 | Western cape | 6 | 0.599 | 0.588 | 3.6 | 3.2 |
| EWC-8 | Western cape | 6 | 0.617 | 0.637 | 3.8 | 3.3 |
| EWC-9 | Western cape | 5 | 0.614 | 0.682 | 3.6 | 3.3 |
| EWC-10 | Western cape | 5 | 0.625 | 0.562 | 3.4 | 3.2 |
| EWCI-11 | Western cape | 5 | 0.642 | 0.577 | 3.6 | 3.3 |
| EWCI-12 | Western cape | 5 | 0.539 | 0.494 | 3.4 | 3.1 |
| EWCI-13 | Western cape | 5 | 0.682 | 0.718 | 3.8 | 3.5 |
| EWCI-14 | Western cape | 6 | 0.596 | 0.539 | 3.4 | 3.0 |
| NWC-1 | Western cape | 4 | 0.668 | 0.632 | 3.5 | 3.5 |
| NWC-2 | Western cape | 5 | 0.656 | 0.588 | 4.0 | 3.6 |
| NWC-3 | Western cape | 4 | 0.591 | 0.618 | 3.2 | 3.2 |
| NWC-4 | Western cape | 5 | 0.676 | 0.635 | 3.9 | 3.6 |
| NWC-5 | Western cape | 12 | 0.644 | 0.598 | 4.8 | 3.4 |
| EC-1 | Eastern Cape | 5 | 0.667 | 0.682 | 3.8 | 3.5 |
| EC-2 | Eastern Cape | 5 | 0.621 | 0.635 | 3.4 | 3.2 |
| EC-3 | Eastern Cape | 5 | 0.617 | 0.553 | 3.6 | 3.4 |
| EC-4 | Eastern Cape | 5 | 0.673 | 0.612 | 3.9 | 3.6 |
| EC-5 | Eastern Cape | 4 | 0.664 | 0.662 | 3.4 | 3.4 |
| ECI-6 | Eastern Cape | 5 | 0.608 | 0.577 | 3.5 | 3.2 |
| ECI-7 | Eastern Cape | 5 | 0.613 | 0.612 | 3.4 | 3.2 |
| ECI-9 | Eastern Cape | 5 | 0.598 | 0.627 | 3.4 | 3.1 |
| ECI-11 | Eastern Cape | 7 | 0.613 | 0.521 | 4.2 | 3.3 |
| ECI-12 | Eastern Cape | 5 | 0.628 | 0.694 | 3.5 | 3.2 |
| ECI-13 | Eastern Cape | 8 | 0.619 | 0.640 | 3.9 | 3.1 |
| NC-1 | Northern cape | 4 | 0.604 | 0.608 | 2.9 | 2.9 |
| FS-1 | Free state | 6 | 0.624 | 0.559 | 3.6 | 3.2 |
| FS-2 | Free state | 5 | 0.631 | 0.624 | 3.6 | 3.3 |
| United Kingdom | | 16 | 0.675 | 0.622 | 6.2 | 4 |

Table S2. Summary of loci characteristics. $H_E$: unbiaised expected heterozygosity; $H_O$: observed heterozyosity; A: number of alleles; and F-statistics ($F_{IS}$, $F_{ST}$, $F_{IT}$).

| | | | South Africa | | United Kingdom | | All populations | | |
|---|---|---|---|---|---|---|---|---|---|
| Locus | Size Range | A | $H_E$ | $H_O$ | He | Ho | $F_{IS}$ | $F_{IT}$ | $F_{ST}$ |
| Ase19 | 156-175 | 6 | 0.562 | 0.472 | 0.635 | 0.500 | 0.173 | 0.180 | 0.008 |
| Patmp2-43 | 124-126 | 2 | 0.441 | 0.436 | 0.370 | 0.333 | 0.036 | 0.046 | 0.011 |
| Pcaμ3 | 172-182 | 5 | 0.715 | 0.665 | 0.669 | 0.563 | 0.098 | 0.112 | 0.016 |
| SS1-6 | 195-213 | 6 | 0.660 | 0.597 | 0.777 | 0.800 | 0.009 | 0.039 | 0.030 |
| SS2-106 | 273-288 | 6 | 0.597 | 0.457 | 0.724 | 0.563 | 0.214 | 0.241 | 0.034 |
| SS2-119 | 274-286 | 6 | 0.618 | 0.597 | 0.502 | 0.500 | 0.006 | 0.040 | 0.034 |
| SS2-130 | 245-255 | 6 | 0.498 | 0.494 | 0.683 | 0.688 | -0.019 | 0.016 | 0.034 |
| SS2-16 | 193-207 | 7 | 0.717 | 0.639 | 0.814 | 0.800 | 0.042 | 0.056 | 0.015 |
| SS2-32 | 228-254 | 13 | 0.880 | 0.818 | 0.832 | 0.533 | 0.197 | 0.232 | 0.044 |
| SS2-68 | 138-140 | 2 | 0.497 | 0.478 | 0.480 | 0.333 | 0.155 | 0.182 | 0.032 |
| SS2-71B | 311-332 | 8 | 0.666 | 0.631 | 0.772 | 0.813 | -0.022 | -0.007 | 0.015 |
| SS2-80 | 306-312 | 2 | 0.279 | 0.275 | 0.480 | 0.333 | 0.182 | 0.223 | 0.051 |
| SS3-42C | 133-161 | 13 | 0.835 | 0.777 | 0.828 | 0.933 | -0.047 | -0.023 | 0.023 |
| Sta213 | 146-193 | 15 | 0.803 | 0.742 | 0.903 | 0.750 | 0.110 | 0.131 | 0.024 |
| Sta269 | 180-203 | 8 | 0.745 | 0.678 | 0.830 | 0.733 | 0.087 | 0.107 | 0.022 |
| Sta294 | 292-305 | 7 | 0.688 | 0.603 | 0.625 | 0.533 | 0.120 | 0.138 | 0.021 |
| Sta308 | 120-158 | 16 | 0.900 | 0.893 | 0.933 | 0.867 | 0.023 | 0.039 | 0.017 |

Figure S1. Sampling sites and historical records of European starling expansion.

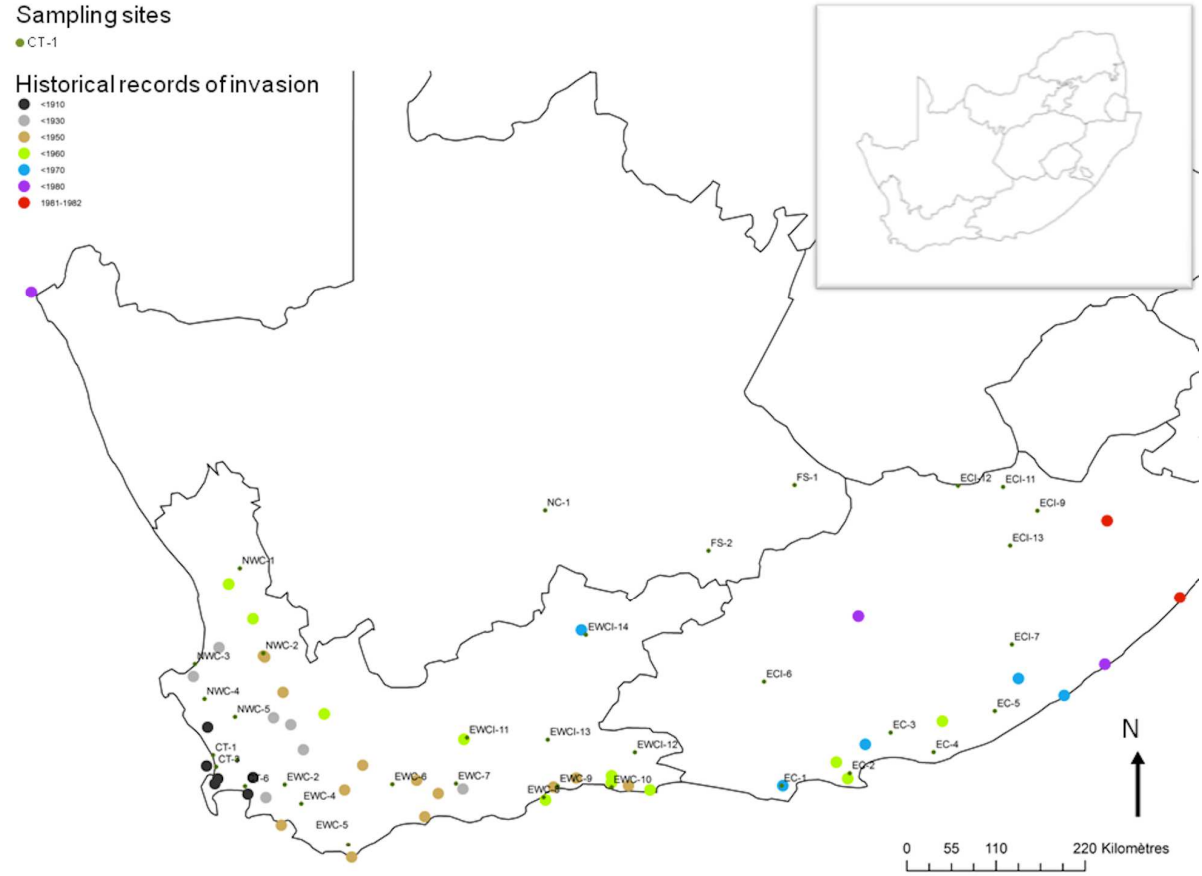



Figure S2. Haplotypes distribution

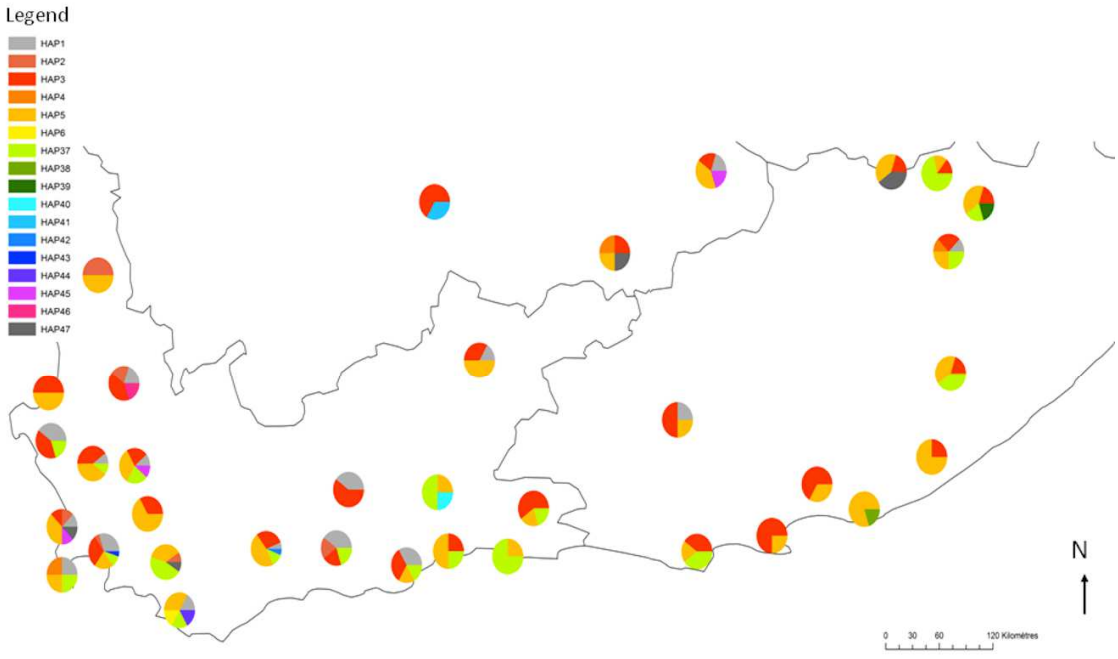


Figure S3. Linear regressions between geographic distance (Ln) and $F_{ST}/(1-F_{ST})$ between sites along the eastern coastal line for sliding windows at the core of range expansion (0-200 km (A); 20-220 km(B); 120-320 km (C)) and at the front of the range expansion ( 820-1020 km (D); 840-1040 km (E); 860-1060 km (F)).

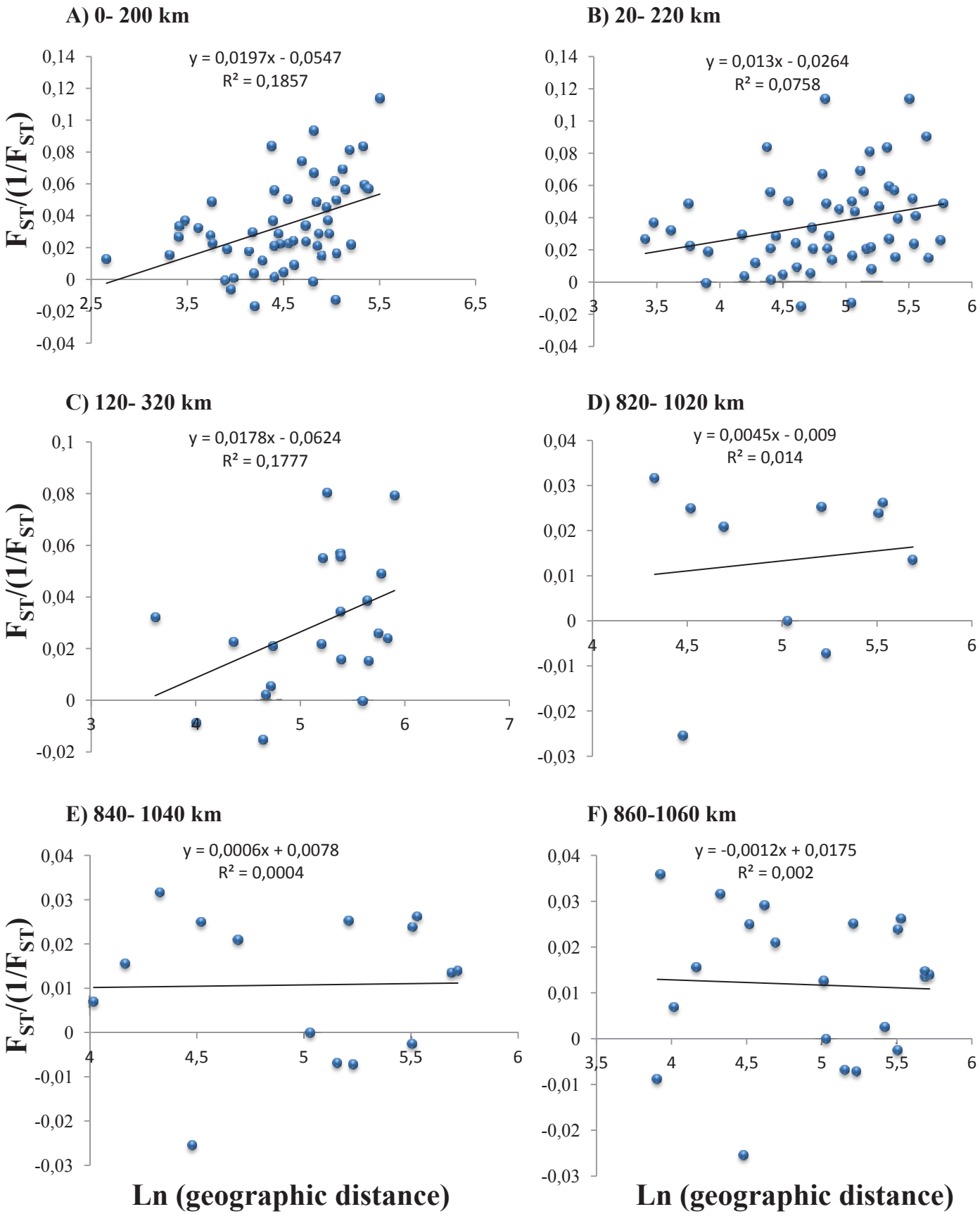
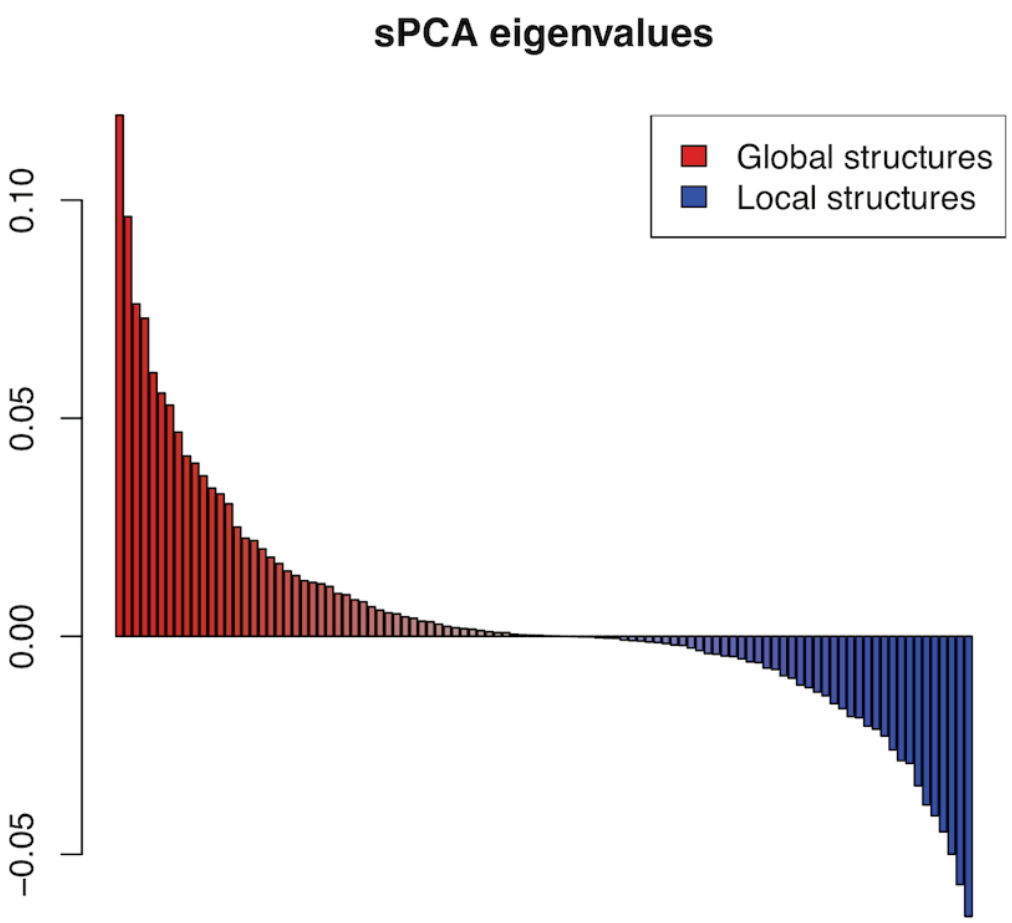


Fig S4. Eigeinvalues of the sPCA



Appendix 1: Bayesian assignment test results from STRUCTURE (Pritchard et al. 2000) analysis: 5 runs for each K value with $10^6$ iterations following a burn-in period of 500,000 using admixture model and correlated allele frequencies were used.

Table 1 from Appendix1. Probability of cluster following Pritchard et al. (2000) method

| K | Probability |
|---|---|
| 1 | 0,16 |
| **2** | **1,00** |
| 3 | 0,27 |
| 4 | 0,11 |
| 5 | 0,07 |

Figure 1 from Appendix1. Log likelihood for each K value

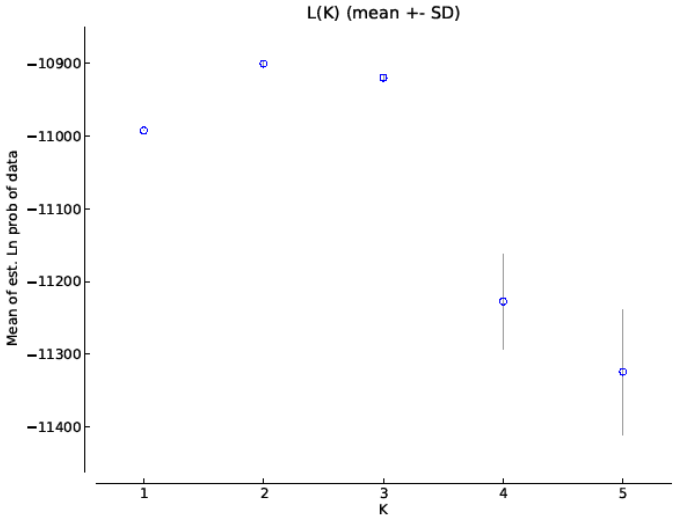



Figure 2 from Appendix1. The inferred number of genetic clusters of the STRUCTURE analysis following the *ΔK* method of Evanno et al. (2005).

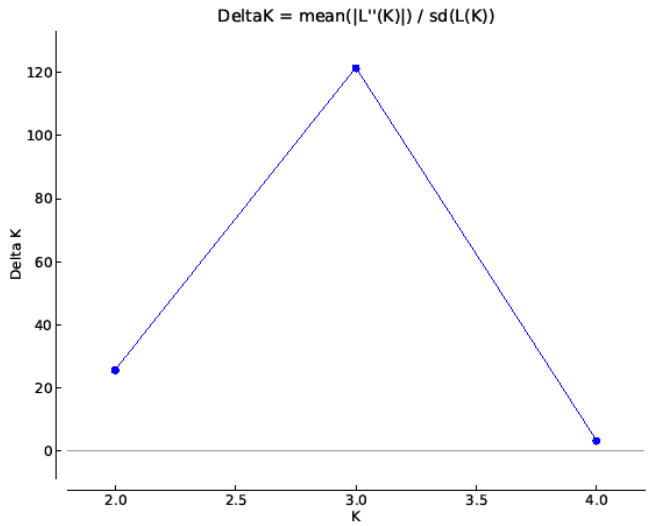


Figure 3 from Appendix1. Mean *q* (assignment) values per sampling site for K=2

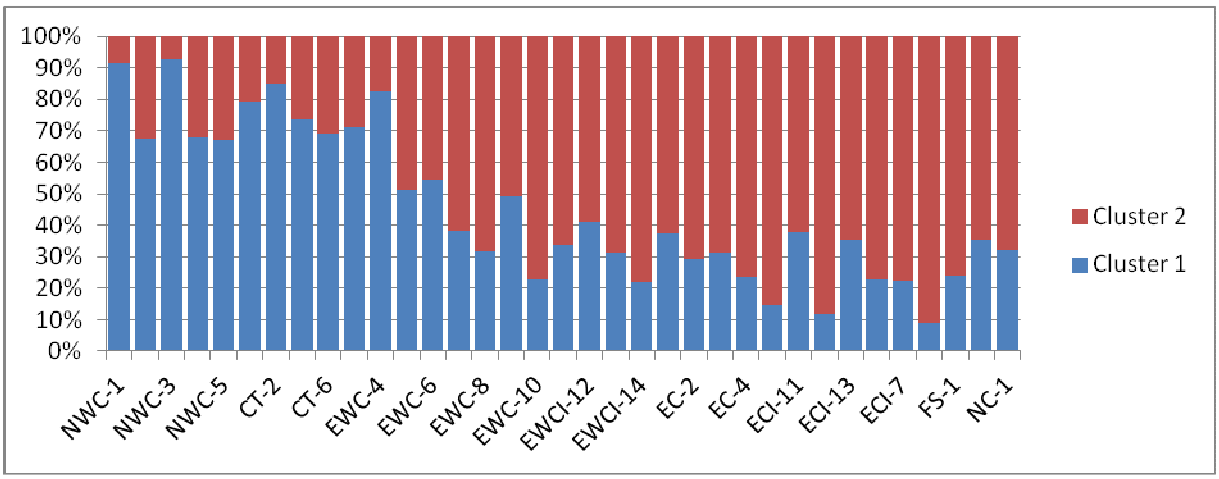



Figure 4 from Appendix1. Mean *q* (assignment) values per sampling site for K=3