INTRODUCTION

One of the most fundamental and oldest objectives of ecology is to understand the factors and mechanisms causing fluctuations in the number of animals in a given area (Elton, 1942; Andrewartha and Birch, 1954; Lack, 1954; Royama, 1992; Coulson et al., 2004). This search for underlying mechanisms has recently acquired a special importance as many natural populations have suffered dramatic declines (Hunter, 2002) because of persecution, exploitation or habitat loss (Beissinger and Snyder, 1992; Casey and Myers, 1998; Krüger et al., 2001, 2010; Ferrer et al., 2003; Pimm et al., 2014).

Traditionally, animal populations have been assessed by means of direct census counts for conservation and management purposes (Luikart et al., 2010). However, an influential contribution of evolutionary theory to conservation biology has been the development of a framework for predicting the fate of small populations (Palstra and Ruzzante, 2008). Central parameters to this framework are the population census size ($N_e$) and effective population size ($N_{ee}$). $N_{ee}$, which is defined as the number of individuals in an ideal population experiencing the same rate of random genetic change over time as the actual population (Crow et al., 1970), is a particularly important quantity as it is inversely proportional to the loss of genetic diversity because of inbreeding and genetic drift in finite, randomly mating populations (Nunney and Elam, 1994; Frankham, 2005; Charlesworth, 2009). Thus, low $N_{ee}$ values are often interpreted as providing an indication of increased extinction risk (Newman and Pilson, 1997).

Although ascertaining $N_{ee}$ is crucial to conservation and management, it is still rather difficult to obtain reliable estimates in real populations because of constraints on collecting enough demographic data to directly measure $N_e$ (Waples, 2005; Luikart et al., 2010). Consequently, molecular genetic approaches have become increasingly popular in recent years (Waples and Do, 2010). A variety of different approaches are now available to estimate $N_{ee}$ from genetic marker data (reviewed by Leberg, 2005; Wang, 2005; Palstra and Ruzzante, 2008; Luikart et al., 2010). These can be broadly classified into single-sample and temporal approaches. The former estimate the effective population size from properties of a single sample of individuals, most often using the unbiased linkage disequilibrium (LD) method of Hill (1981) that is based on the premise that in small populations with few parents, random genetic drift generates nonrandom associations between alleles at different loci. In contrast, temporal approaches exploit the fact that drift is higher in smaller populations by quantifying differences in allele frequencies between two or more samples collected at different time points. Applied to discrete cohorts of offspring, these approaches strictly estimate the effective number of parents that produced the sample, or $N_b$ (Schwartz et al., 1999; Leberg, 2005; Waples and Do, 2010; Waples and England, 2011; Waples et al., 2014).

Both sets of approaches have advantages and disadvantages. Temporal approaches are often considered to be the most accurate, assuming that samples can be gathered sufficiently far apart in time (Leberg, 2005; Luikart et al., 2010). However, they require at least two non-overlapping generations to be sampled, and this is often impossible for species with long generation times such as many large vertebrates (Luikart et al., 2010). Single-sample approaches have therefore increased in popularity as they can provide a snapshot of a population without the need to collect multiple samples. Precision also appears to be reasonably high, at least for the LD method, when using the 10–20 genetic markers and ~ 50 samples typical of most studies, as

---

**ORIGINAL ARTICLE**

Long-term effective population size dynamics of an intensively monitored vertebrate population

A-K Mueller, N Chakarov2, O Krüger and JI Hoffman

Long-term genetic data from intensively monitored natural populations are important for understanding how effective population sizes ($N_{ee}$) can vary over time. We therefore genotyped 1622 common buzzard (Buteo buteo) chicks sampled over 12 consecutive years (2002–2013 inclusive) at 15 microsatellite loci. This data set allowed us to both compare single-sample with temporal approaches and explore temporal patterns in the effective number of parents that produced each cohort in relation to the observed population dynamics. We found reasonable consistency between linkage disequilibrium-based single-sample and temporal estimators, particularly during the latter half of the study, but no clear relationship between annual $N_{ee}$ estimates ($\tilde{N}_{ee}$) and census sizes. We also documented a 14-fold increase in $N_{ee}$ between 2008 and 2011, a period during which the census size doubled, probably reflecting a combination of higher adult survival and immigration from further afield. Our study thus reveals appreciable temporal heterogeneity in the effective population size of a natural vertebrate population, confirms the need for long-term studies and cautions against drawing conclusions from a single sample.

*Heredity (2016) 117, 290–299; doi:10.1038/hdy.2016.67; published online 24 August 2016*
long as $N_e$ is less than $\sim 500$ (Waples and Do, 2010). However, single-sample approaches can be sensitive to gene flow (Gilbert and Whitlock, 2015) and there is also a need for studies that are long enough to quantify temporal variability when applying temporal methods (Palstra and Ruzzante, 2008).

A long-term study of common buzzards (Buteo buteo) provides an excellent opportunity to explore the temporal dynamics of the effective population size of an intensively studied vertebrate population. This species is a common, medium-sized bird of prey that breeds across the entire Palaearctic from southwestern Europe to Japan, where it preys mainly on microtine rodents (del Hoyo et al., 1994). A population in Eastern Westphalia, Germany, has been intensively monitored since 1989, with nests having been climbed and all chicks marked to allow individual recognition since 2002. Our census data suggest that this population was relatively stable at $\sim 200$ breeding adults since the early 2000s, but began to increase in 2010 towards a peak of 516 in 2012 (see Results).

Common buzzards also exhibit a striking plumage polymorphism (Cramp and Simmons, 1980; Krüger et al., 2001) that serves as a phenotypic marker and thus provides an interesting additional dimension (Chakarov et al., 2008; Chakarov et al., 2013). Three colour morphs have been described that differ in their levels of plumage melanisation, termed ‘light’, ‘intermediate’ and ‘dark’. Plumage morph is fixed throughout an individual’s lifetime, follows Mendelian expectations for a single locus with two alleles (Krüger et al., 2001) and is associated with a mutation in the melanocortin-1 receptor (MA Pointer et al., unpublished data). The intermediately melanised morph, which is presumed to be heterozygous at the colour locus, is on average longer lived and has the greatest lifetime reproductive success (Chakarov et al., 2008). However, the plumage morphs do not differ significantly in genome-wide heterozygosity measured using 18 microsatellites (Boerner et al., 2013), suggesting that heterozygote advantage is not a genome-wide phenomenon. The three morphs are also genetically undifferentiated, despite buzzards tending to mate with partners of the same morph as their mother (Boerner et al., 2013).

Here, we genotyped 1622 common buzzard chicks comprising 12 complete cohorts at 15 polymorphic microsatellite loci. The resulting data were used to evaluate the comparability of single-sample and temporal estimators, as well as to explore temporal patterns in relation to the observed demography of the population.

**MATERIALS AND METHODS**

**Study site**

This study was conducted in a 300 km$^2$ study area (8°25'E and 52°6’N) in Eastern Westphalia, Germany (Figure 1). It consists of two 125 km$^2$ grid squares and 50 km$^2$ of edge areas. The dominant land cover in the study area is the Teutoburger Wald, a low-level forested mountain region reaching a height

---

**Figure 1** Location of the study area in Germany together with a detailed map (insert) of the 300 km$^2$ study area, with human settlements shown in orange, forest patches in green and agricultural areas in cream. A full colour version of this figure is available at the *Heredity* journal online.
of 315 m above sea level. The second most abundant land cover type is a cultivated landscape to the north and south.

Collection of census and breeding success data

Census dynamics of common buzzards were monitored from 1989 to 2013 inclusive. All forest patches were visited in late winter to look for territorial pairs and nests. During the breeding season, these forest patches were visited again and checked for activity of the study species. This includes breeding pairs (occupying a nest and showing signs of egg-laying activity) as well as nonbreeding pairs that just occupy a territory. Hence, our buzzard census data include breeding and nonbreeding pairs but the number of ‘floating’ individuals that do not hold a territory cannot be reliably counted and recorded.

Each active nest was visited at least 3 and up to 10 times a year to determine breeding success (success or failure) and brood size (number of chicks fledged) for successful breeding attempts. From 1989 to 2001 inclusive, data were collected through careful and intensive observation from the ground. In subsequent years, between 85 and 99% of all successful nests were climbed and the chicks were ringed, normally in late May and early June.

Blood sampling

Buzzard nests were climbed with a rope-climbing technique and once the climber was at the nest, chicks were lowered to the ground using another rope. On the ground, they were ringed, biometric measures were recorded and a 0.5 ml blood sample was taken from the brachial vein with a syringe or needles and capillaries. Sample sizes are shown in Table 1. Blood was transferred into 1.5 ml screw-cap tubes filled with 1.0 ml ethanol or phosphate-buffered saline–EDTA buffer. Back at the laboratory, all tubes were stored at ~ 20 °C.

Microsatellite genotyping

Total genomic DNA was extracted from 10–20 μl of each sample using a standard chloroform extraction protocol and genotyped at 15 previously developed microsatellite loci (Johnson et al., 2005). All but two of these loci map to different loci in the zebra finch (T. guttata) and are therefore unlikely to be physically linked (Table 2). The microsatellites were PCR amplified in a single multiplexed reaction using a Type It Kit (Qiagen GmbH, Hilden, NW, Germany). The following PCR profile was used: one cycle of 5 min at 94 °C; 24 cycles of 30 s at 94 °C, 90 s at 56 °C and 30 s at 72 °C; and one final cycle of 15 min at 72 °C. Fluorescently labelled PCR products were then resolved by electrophoresis on an ABI 3730xl capillary sequencer (Applied Biosystems, Carlsbad, CA, USA) and allele sizes were scored automatically using GeneMarker version 2.6.2 (Softgenetics, AB).

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Sample number</th>
<th>upper %</th>
<th>lower %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>109</td>
<td>140</td>
<td>109</td>
</tr>
<tr>
<td>2002</td>
<td>79</td>
<td>110</td>
<td>79</td>
</tr>
<tr>
<td>2001</td>
<td>502</td>
<td>208</td>
<td>502</td>
</tr>
<tr>
<td>2000</td>
<td>83</td>
<td>164</td>
<td>83</td>
</tr>
<tr>
<td>2004</td>
<td>38</td>
<td>66</td>
<td>38</td>
</tr>
<tr>
<td>2009</td>
<td>138</td>
<td>182</td>
<td>138</td>
</tr>
<tr>
<td>2010</td>
<td>298</td>
<td>228</td>
<td>298</td>
</tr>
<tr>
<td>2011</td>
<td>262</td>
<td>220</td>
<td>262</td>
</tr>
<tr>
<td>2012</td>
<td>126</td>
<td>148</td>
<td>126</td>
</tr>
</tbody>
</table>

Intermediate (739) 268 739

Light (502) 208 502

Table 1 Sample sizes of common buzzards with the number of analysed samples (N), census population size ($N_c$), number of breeders ($N_b$), estimated effective number of parents that produced each cohort ($N_e$), minimal allele frequency cutoff ($P_{cutoff}$) and 95% confidence intervals (CIs)

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Sample number</th>
<th>upper %</th>
<th>lower %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>109</td>
<td>140</td>
<td>109</td>
</tr>
<tr>
<td>2002</td>
<td>79</td>
<td>110</td>
<td>79</td>
</tr>
<tr>
<td>2001</td>
<td>502</td>
<td>208</td>
<td>502</td>
</tr>
<tr>
<td>2000</td>
<td>83</td>
<td>164</td>
<td>83</td>
</tr>
<tr>
<td>2004</td>
<td>38</td>
<td>66</td>
<td>38</td>
</tr>
<tr>
<td>2009</td>
<td>138</td>
<td>182</td>
<td>138</td>
</tr>
<tr>
<td>2010</td>
<td>298</td>
<td>228</td>
<td>298</td>
</tr>
<tr>
<td>2011</td>
<td>262</td>
<td>220</td>
<td>262</td>
</tr>
<tr>
<td>2012</td>
<td>126</td>
<td>148</td>
<td>126</td>
</tr>
</tbody>
</table>

Long-term effective population size dynamics

A-K Mueller et al.

analyses of molecular variance within GenAlEx version 6.5 (Peakall and Smouse, 2012) to test for genetic differences between buzzards sampled to the north and south of the ridge of the Teutoburger Wald.

Wild populations were further studied using hierarchical genetic data analyses of molecular variance within GenAlEx version 6.5 (Peakall and Smouse, 2012) to test for genetic differences between buzzards sampled to the north and south of the ridge of the Teutoburger Wald.

Wild populations were further studied using hierarchical genetic data analyses of molecular variance within GenAlEx version 6.5 (Peakall and Smouse, 2012) to test for genetic differences between buzzards sampled to the north and south of the ridge of the Teutoburger Wald.
To explore sensitivity to the number of generations assumed to
estimate is 4.75 years. For the temporal analysis, we therefore took the allele
(CIs) were derived using the
parametric
option that implements
χ² approximation (Waples, 2006).

Temporal estimators
We used the software NeEstimator version 2.01 (Do et al., 2014) to implement
the LD approach of Waples and Do (2008). As the inclusion of rare alleles can
upwardly bias LD-based estimates (Waples and Do, 2010), we followed the
authors’ recommendation of choosing the minor allele frequency threshold (Pcrit)
to be the larger of 0.02 or a value that screens out single copy alleles
(Waples and Do, 2010). To explore sensitivity, we also repeated the analysis
with Pcrit = 0.01 and 0.05 while similarly applying the same criterion as above to
ensure that single copy alleles were not counted. The 95% confidence intervals (CIs)
were derived using the ‘parametric’ option that implements χ² approxima-
tion (Waples, 2006).

RESULTS
We genotyped 1622 common buzzard individuals sampled over
12 years at 15 microsatellite loci. Individuals that failed to produce
interpretable genotypes at three or more loci were discarded, leaving a
total of 1419 individuals for the data analysis (Table 1). The loci were
moderately variable, carrying an average of 9.75 alleles (Table 2). Tests
for deviations of each locus from Hardy–Weinberg equilibrium in each
of the years revealed the number of deviations that remained significant
following table-wide false discovery rate correction for multiple tests
(Supplementary Table S1). Two of the loci (Bbu 35 and 46) deviated significantly from Hardy–Weinberg equilibrium in 9 out of
12 years and a further locus (Bbu 26) deviated significantly in
7 years. We therefore took the conservative measure of excluding all
three of these loci from further analyses. Tests for LD among the
remaining 12 loci revealed a number of significant associations
(Figure 2). However, no pairs of loci were consistently in LD across
multiple years, the extent of LD varied from year to year (Figure 2)
and all but two of the loci mapped to different chromosomes in the
zebra finch (Table 2), suggesting that these associations are unlikely to
be due to physical linkage.

B Buzzard morph effective population sizes
To test for differences in the effective population sizes of the
three buzzard plumage morphs, we implemented the LD method
(Hill, 1981; Waples, 2006; Waples and Do, 2010). For this analysis, it
was necessary to pool individuals across years because of the low
frequency of the dark morph. Sensitivity to the minor allele frequency
cutoff was assessed by generating effective population size estimates
for Pcrit = 0.01, 0.02 and 0.05. The resulting estimates, which were
reasonably robust to the Pcrit value used, were lowest for the dark
morph, intermediate for the light morph and highest for the
intermediate morph (Figure 3 and Table 1), reflecting their frequen-
cies in the wider population.

Table 2 Details of the 12 microsatellite loci used in this study

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome in Taeniopygia guttata</th>
<th>Allele number</th>
<th>Observed heterozygosity</th>
<th>Expected heterozygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bbu03</td>
<td>15 8</td>
<td>0.70</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>Bbu06</td>
<td>5 7</td>
<td>0.50</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Bbu11</td>
<td>5 10</td>
<td>0.49</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Bbu14</td>
<td>7 12</td>
<td>0.55</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>Bbu17</td>
<td>1 9</td>
<td>0.55</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>Bbu22</td>
<td>1 4</td>
<td>0.13</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Bbu30</td>
<td>2 5</td>
<td>0.23</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Bbu33</td>
<td>3 10</td>
<td>0.79</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Bbu34</td>
<td>1 17</td>
<td>0.81</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>Bbu38</td>
<td>2 5</td>
<td>0.11</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Bbu42</td>
<td>4 25</td>
<td>0.87</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Bbu53</td>
<td>19 4</td>
<td>0.23</td>
<td>0.26</td>
<td></td>
</tr>
</tbody>
</table>

Single-sample estimators
We used the software NeEstimator version 2.01 (Do et al., 2014) to implement
the LD approach of Waples and Do (2008). As the inclusion of rare alleles can
upwardly bias LD-based estimates (Waples and Do, 2010), we followed the
authors’ recommendation of choosing the minor allele frequency threshold (Pcrit)
to be the larger of 0.02 or a value that screens out single copy alleles
(Waples and Do, 2010). To explore sensitivity, we also repeated the analysis
with Pcrit = 0.01 and 0.05 while similarly applying the same criterion as above to
ensure that single copy alleles were not counted. The 95% confidence intervals (CIs)
were derived using the ‘parametric’ option that implements χ² approxima-
tion (Waples, 2006).

Temporal estimators
We used the temporal approach of Wang and Whitlock (2003) as this is the
least sensitive of the temporal estimators to immigration (Gilbert and Whitlock,
2015). The software MLNe 1.0 (Wang and Whitlock, 2003) was used to
generate both the maximum-likelihood and moment estimators, assuming that
the population is not at equilibrium. For this analysis, we specified a maximum
Nc of 5000, a monitor value of four and six threads. We estimated the
generation time as the mean age at maturity plus the mean reproductive
lifespan (IUCN Standards and Petitions Subcommittee, 2016). As the breeding
lifespan of an adult female buzzard is on average 2.75 years and most buzzards
recruit as breeding adults at ∼2 years of age (Krüger and Lindström, 2001), our
estimate is 4.75 years. For the temporal analysis, we therefore took the allele
frequencies from 2002 and 2013 and assumed two generations between
samples. To explore sensitivity to the number of generations assumed to
separate the samples, we also repeated this analysis specifying between one and
five generations.

Population structure
As population structure can cause inaccuracies in Ne estimation
(Waples and England, 2011; Neel et al., 2013; Gilbert and Whitlock,
2015), we used two approaches to test for the presence of cryptic
population structure within the study area. Arguably the most versatile
tests of population structure need not rely on knowledge of where
individuals were sampled. Consequently, we first implemented a

Long-term effective population size dynamics
A-K Mueller et al
Bayesian cluster analysis using the program STRUCTURE (Pritchard et al., 2000) to determine whether any genetic structure could be detected in the absence of a priori geographic data. The resulting posterior probabilities were highly concordant among replicate runs, with the highest average value indicating the most likely number of clusters, $K$. The average log likelihood value climbed steadily with increasing $K$ to peak at $K=20$ (Supplementary Figure S3). However, when this analysis was repeated separately for each of the twelve years, the maximal average log likelihood values were mainly associated with $K=1$, suggesting a lack of detectable population structure within years (Supplementary Figure S4). The only exceptions were 2005 and 2013, for which the most likely genetic structure consisted of six and four
clusters respectively. Visual inspection of individual cluster memberships for these years revealed considerable admixture and no clear evidence for the presence of distinct sub-populations.

As a further test of population structure that exploits prior information on sampling locations, we implemented an analysis of molecular variance. For this analysis, we compared animals sampled to the north and south of the Teutoburger Wald, a low mountain region that bisects the study area. Around 1% of the variance was partitioned between the north and south ($F = 0.03$, $P = 0.001$) indicating the presence of weak but statistically significant population substructure. To test whether this has an effect on the effective population size estimates, we repeated the temporal analysis shown in Figure 5a separately for the northern and southern sub-populations. The same overall trend was observed for the northern sub-population, whereas $N_e$ values were both lower and less variable over time for the southern sub-population (Supplementary Figure S1A). This is consistent with census data (Supplementary Figure S1B) and suggests that the overall pattern shown in Supplementary Figure S1 is driven by changes in the larger northern sub-population.

**Sensitivity to sample size**

Sample sizes were generally lower in the early part of the study, which appears to be reflected to some extent in the corresponding effective population size estimates. We therefore tested for a relationship between annual $N_e$ and sample size. This was not statistically significant ($F_{1,10} = 2.76$, $P = 0.13$), suggesting that sample size does not have a strong effect on the magnitude of the estimates. To further explore whether our conclusions could be affected by variation in sample size, we generated $N_e$ values for pooled data corresponding to the periods 2002–2009 and 2010–2013 inclusive. To mimic sampling effects, we then randomly selected differently sized subsets of individuals each 10 times. The resulting $N_e$ values were consistently greater for the latter period (Supplementary Figure S2) and no clear relationship was found between $N_e$ and sample size for either period.

**DISCUSSION**

Long-term genetic studies are essential for understanding temporal variation in the effective population size of natural populations. We therefore generated a large microsatellite data set for an intensively monitored common buzzard population in northern Germany. The 12-year duration of the study allowed us to use both single-sample and temporal approaches that yielded comparable estimates, at least for the latter part of the study. Analysis of 12 successive cohorts also uncovered appreciable temporal heterogeneity, with $N_e$ varying by a factor of 14 over the course of the study.

We used multiple approaches to estimate the effective population size of our study population of common buzzards. However, because of the fact that the population has expanded in recent years, we took the conservative measure of focussing on those estimators that are least sensitive to immigration into the focal population (Gilbert and Whitlock, 2015). We found that the results of the LD method were reasonably consistent across a range of $P_{crit}$ values from 0.01 to 0.05. One relevant feature of common buzzards is the presence of three distinct colour morphs that differ markedly in their frequency in the study population (Krüger et al., 2001). Accordingly, $N_e$ was smallest for the dark morph, intermediate for the light morph and largest for the intermediate morph. Although heterozygote advantage operates at the colour locus in this population, with the intermediate morph having the greatest average lifetime reproductive success (Chakarov et al., 2008), we recently showed that the three morphs do not differ significantly in their genome-wide heterozygosity, nor are they genetically differentiated from one another (Boerner et al., 2013). Variation in the effective population sizes of the morphs is therefore unlikely to be an artefact of differences in genome-wide heterozygosity or population substructure and instead appears to be a reflection of the relative frequencies of the morphs in the population.

We also used the LD method to generate annual effective population size estimates. These generally had rather small 95% confidence intervals, consistent with a simulation study (Waples and Do, 2008) showing that the LD method can generate precise estimates with ~10–20 microsatellites and 50 individuals sampled where $N_e$ is less than ~500. The estimates themselves varied by a factor of 14 over the duration of the study and were consistently lower during 2002–2009. The smallest estimate of 25.5 was obtained for 2008, a year in which breeding success was particularly low. Afterwards, the estimates steadily increased towards a peak of 367.2 in 2011. Although the exact causes of this increase are not known, census data indicate that the population more than doubled from 2009 to 2012, whereas over the same period the frequency of the white morph increased from 30 to 41%. Hence, temporal patterns in $N_e$ appear to capture underlying population processes that are also reflected in the census data.
Possible causes for population growth include increased survival and local recruitment, and immigration from further afield. However, the former is unlikely to have played a major role as, despite having fitted almost all fledged chicks from the study area with clearly visible wing tags, we have not observed an increase in the number of local birds recruiting into the population (Chakarov et al., 2013). In fact, many new territories have been found by individuals without wing tags (O Krüger, personal communication). As we are confident that these individuals are not local, the only explanation remaining is that population growth is attributable at least in part to immigration. This is also consistent with changes over time both in the frequency of the light morph and in the proportion of pairwise comparisons among loci yielding significant LD values (see below).

As the LD approach will be affected by the presence of physically linked markers, we checked for LD among the 12 microsatellites in each of the years. We did not find consistent patterns of association between particular pairs of loci, and this agrees with the fact that most of the markers map to different chromosomes in the zebra finch. However, the proportion of pairwise comparisons yielding significant test statistics varied from year to year, being mostly low but reaching highs of $\sim 53\%$ in 2010 and 38\% in 2012 (Figure 2). Although this could partly reflect larger sample sizes for these two particular years, over the same period we found very little variation in either observed heterozygosity (Table 2) or standardised allelic richness (Table 3), suggesting that genetic diversity has not altered appreciably. Similarly, the results of the Structure analysis suggest that there has been no apparent increase in the amount of population structure over time in parallel with the $N_e$ estimates. Thus, it seems likely that NeEstimator is detecting changes in LD within the population. These changes could potentially be a consequence of gene flow into the focal population, as...
this is known to create LD when allele frequencies are unequal among populations exchanging migrants (Gilbert and Whitlock, 2015). Although it has been estimated that 80% of buzzards recruit to within 20 km of their natal territories (Zang et al., 1989) and 96% within 100 km (Walls and Kenward, 1998), our genetic data are not consistent with localised immigration from the vicinity of our study population, as the high mobility of common buzzards should reduce population structure over this scale. However, longer-distance dispersal does happen regularly in this species (Cramp and Simmons, 1980; Kenward et al., 2001) and buzzard dispersal is strongly influenced by weather patterns (Walls et al., 2005). In particular, harsh winters are known to severely affect buzzards (Cramp and Simmons, 1980) and the winter of 2009–2010 was among the harshest in recent years that might have induced significant dispersal events. On current evidence, it remains speculative where the influx of light birds may have come from, as there is no clear evidence for cline variation in buzzard morph frequency.

One caveat to the LD method is that LD can be generated by many different phenomena, from inbreeding through population structure to immigration (Luikart et al., 2010). Inbreeding is unlikely to be important in this system, partly because the species is monogamous but also because a previous study found no evidence for inbreeding (Boerner et al., 2013). However, population structure could potentially be present as the study area is bisected by the Teutoburger Wald. To check for the presence of discrete populations within the study area, we therefore used Bayesian cluster analysis and analysis of molecular variance. The Structure results hinted at the presence of multiple clusters when all of the data were analysed together but generally indicated a lack of structure when the years were analysed separately. Such a pattern could potentially arise because of the fact that buzzard pairs often breed across multiple years, meaning that a stronger signal of family structure may be present in the full data set relative to individual years.

In contrast, analysis of molecular variance uncovered a small but significant genetic difference between buzzards breeding to the north and south of the Teutoburger Wald. This most likely reflects differences in habitat suitability, as similar differences between the northern and southern parts of our study area have been documented for buzzard survival in interaction with local weather patterns (Jonker et al., 2014). This is probably a reflection of the predominance of scotts pine forest on sandy soils in the southern area that is known to be a suboptimal habitat for buzzards (Krüger, 2004).

Sensitivity of effective population size estimators to violation of the assumption of discrete, non-overlapping generations is an issue that hampers many studies of natural populations (Waples et al., 2014, Kamath et al., 2015). Although samples that combine multiple cohorts (strictly, as many cohorts as there are in a generation) estimate $N_e$, single-cohort samples are usually thought of as providing information about the effective number of parents that produced the sample, that is, $N_0$ (Schwartz et al., 1999; Leberg, 2005; Waples and Do, 2010; Waples and England, 2011; Waples et al., 2014). With knowledge of the ratio of $N_0/N_e$, which can be estimated from life history information (Waples et al., 2013), it is possible to directly estimate $N_e$. However, this was not possible for our study population because the majority of breeding individuals are of unknown age.

Another important caveat is that for the temporal method the number of generations between samples affects both precision and bias (Waples and Yokota, 2007; Gilbert and Whitlock, 2015). Precision can be reasonably good with as little as two generations separating samples, but age structure can cause bias by violating the assumption of discrete generations (Jorde and Ryman, 1995; Wang et al., 2010). Although the magnitude and direction of bias depends on the species’ life history and the type of samples taken, as a guideline it has been suggested that at least 3–5 generations are needed to minimise bias (Waples and Yokota, 2007). In the case of our study population, wing-tag data indicate that local recruitment is negligible, suggesting that there will be little if any overlap between individuals sampled as chicks at the beginning of the study and adults that contributed chicks towards the population in the later part of the study. However, to quantify the magnitude of bias will require at least one more decade of genetic sampling.

Despite the above caveats, we believe that our main conclusions are unlikely to be strongly affected by the presence of overlapping generations for three main reasons. First, we are more interested in the relative values of the estimates rather than the absolute values. This is why we estimated $N_e$ based on pooled samples for the different colour morphs, as the dark morph in particular is too infrequent to be able to generate meaningful estimates separately for each of the years. Second, although the results of the temporal analysis may be subject to bias (see above) and vary with the precise estimator used and the number of generations assumed to separate the samples, arguably the temporal estimates are reasonably consistent with the single-sample $N_e$ estimates obtained for the latter years of the study, as also shown by Miller and Waits (2003) and Rowe and Beebee (2004). Third, any temporal variation in $N_e$ across the study will tend to be dampened by individuals who contribute towards estimates in successive years, making our temporal analysis if anything somewhat conservative.

One potential issue with our study is that sample sizes varied from year to year and were generally lower in the first half of the study where the corresponding $N_e$ values were also smaller. We therefore tested whether the conclusion of a temporal increase in $N_e$ was robust to sample size variation. For this analysis, we pooled data for the periods 2002–2009 and 2010–2013 inclusive, and then calculated $N_e$ for each period using differently sized random subsamples. Regardless of sample size, consistently larger estimates were obtained for the latter period, in support of a genuine temporal increase in $N_e$. Consistent with the observation that annual $N_e$ did not correlate significantly with sample size, we also found no indication of an increase in $N_e$ with the number of randomly selected individuals.

A handful of previous studies have used similar approaches to explore temporal patterns in the effective population size of natural populations, although for species like bears with relatively long generation times this requires the analysis of museum specimens.

Table 3 Standardised allelic richness with s.d. for the 12 microsatellite loci used in this study for each year

<table>
<thead>
<tr>
<th>Years</th>
<th>Allelic richness</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>5.25</td>
<td>0.236</td>
</tr>
<tr>
<td>2003</td>
<td>4.75</td>
<td>0.236</td>
</tr>
<tr>
<td>2004</td>
<td>4.73</td>
<td>0.199</td>
</tr>
<tr>
<td>2005</td>
<td>4.92</td>
<td>0.312</td>
</tr>
<tr>
<td>2006</td>
<td>5.08</td>
<td>0.000</td>
</tr>
<tr>
<td>2007</td>
<td>5.17</td>
<td>0.349</td>
</tr>
<tr>
<td>2008</td>
<td>4.95</td>
<td>0.24</td>
</tr>
<tr>
<td>2009</td>
<td>5.05</td>
<td>0.385</td>
</tr>
<tr>
<td>2010</td>
<td>4.63</td>
<td>0.336</td>
</tr>
<tr>
<td>2011</td>
<td>4.87</td>
<td>0.326</td>
</tr>
<tr>
<td>2012</td>
<td>5.25</td>
<td>0.317</td>
</tr>
<tr>
<td>2013</td>
<td>4.58</td>
<td>0.257</td>
</tr>
</tbody>
</table>


Jonker RM, Chakarov N, Krüger O (2014). Climate change and habitat heterogeneity drive a population increase in Common Buzzards *Buteo buteo* through effects on survival. *Ibis* 156: 97–106.


