Negative effects of individual heterozygosity on reproductive success in a wild bird population

Esteban Botero-Delgadillo | Carol Gilsenan | Jakob C. Mueller | Bart Kempenaers

Abstract
The evolutionary consequences of individual genetic diversity are frequently studied by assessing heterozygosity–fitness correlations (HFCs). The prevalence of positive and negative HFCs and the predominance of general versus local effects in wild populations are far from understood, partly because comprehensive studies testing for both inbreeding and outbreeding depression are lacking. We studied a genetically diverse population of blue tits in southern Germany using a genome-wide set of 87 microsatellites to investigate the relationship between proxies of reproductive success and measures of multilocus and single-locus individual heterozygosity (MLH and SLH). We used complimentary measures of MLH and partitioned markers into functional categories according to their position in the blue tit genome. HFCs based on MLH were consistently negative for functional loci, whereas correlations were rather inconsistent for loci found in nonfunctional areas of the genome. Clutch size was the only reproductive variable showing a general effect. We found evidence for local effects for three measures of reproductive success: arrival date at the breeding site, the probability of breeding at the study site and male reproductive success. For these, we observed consistent, and relatively strong, negative effects at one functional locus. Remarkably, this marker had a similar effect in another blue tit population from Austria (~400 km to the east). We suggest that a genetic local effect on timing of arrival might be responsible for most negative HFCs detected, with carry-over effects on other reproductive traits. This effect could reflect individual differences in the distance between overwintering areas and breeding sites.

Keywords
heterozygosity–fitness correlations, homozygote advantage, microsatellites, outbreeding depression, single-locus effects
1 | INTRODUCTION

Investigating the genetic basis of variation in fitness is essential to increase our understanding of evolution in wild populations. For decades, researchers have aimed to assess the effects of mating with close relatives and of outcrossing between populations (Lynch & Walsh, 1998). Empirical evidence suggests that, in general, inbreeding within populations can cause a reduction in mean fitness, that is inbreeding depression, whereas outbreeding commonly has a positive effect on fitness-related traits (Charlesworth & Willis, 2009; Lynch, 1991). However, population admixture can also negatively impact fitness by means of intrinsic genetic costs or due to the dilution of locally adapted genomes, that is outbreeding depression (Lynch, 1991; Verhoeven, Macel, Wolfe, & Biere, 2010). It is known that inbreeding and outbreeding depression occur in nature, but the prevalence of one over the other in wild populations requires further examination (Szulkin & David, 2011). As a result, several studies have focused on assessing the effect of genetic diversity at the individual level on life-history, morphological and physiological traits (see review in Chapman, Nakagawa, Colman, Slate, & Sheldon, 2009).

One of the most popular approaches to investigate the evolutionary consequences of individual genetic diversity consists of correlating measures of heterozygosity with fitness components, also known as heterozygosity–fitness correlations (hereafter HFCs; reviewed in Chapman et al., 2009; Colman & Slate, 2003; Szulkin, Bierne, & David, 2010). According to heterozygosity theory, HFCs inferred from selectively neutral markers can be interpreted to be a result of a genome-wide effect—“general effect hypothesis”—or a localized, single-locus effect—“local effect hypothesis”—(for a review, see Kempenaers, 2007). In the former case, differences in average fitness between heterozygotes and homozygotes result from a genome-wide association between a group of markers and fitness loci. For this to happen, heterozygosity/homozygosity should covary across loci—that is identity disequilibrium (ID; David, 1998)—so that a multilocus measure of heterozygosity (MLH) will be representative of genome-wide inbreeding/outbreeding (Kempenaers, 2007; Szulkin et al., 2010). In the latter case, an effect on fitness is a product of the association of a particular marker with a functional locus, either due to physical linkage or due to linkage disequilibrium—gametic phase disequilibrium (LD; Lewontin & Kojima, 1960). A direct effect of the marker is also possible (see “direct effect hypothesis”; Kempenaers, 2007), but this will only occur if the functional locus itself has been scored (e.g. allosymes or MHC loci; reviewed in David, 1998).

Given their relevance in evolutionary biology and conservation genetics, HFC studies have been carried out in several taxa and reviewed in detail by numerous authors (e.g. Chapman et al., 2009; Kardos, Allendorf, & Luikart, 2014; Kempenaers, 2007; Miller & Colman, 2014; Mueller, Hermisson, Olano-Marin, Hansson, & Kempenaers, 2011). In general, the biological meaning and generality of HFCs have been questioned for different reasons that include the presumed limited capacity of heterozygosity indexes to measure individual internal relatedness (Balloux, Amos, & Coulson, 2004; but see Forstmeier, Schielzeth, Mueller, Ellegren, & Kempenaers, 2012), publication bias favouring significant results (Colman & Slate, 2003) and the lack of statistical power to detect correlations (Kardos et al., 2014; Miller & Colman, 2014). In addition, results appear to be equivocal, as different studies have reported positive, negative or no HFCs (e.g. Boerner, Hoffman, Amos, Chakarov, & Kruger, 2013; Grueber, Laws, Nakagawa, & Jamieson, 2010; Judson, Knapp, & Welch, 2018; Lieutenant-Gosselin & Bernatchez, 2006; Olano-Marin, Mueller, & Kempenaers, 2011a; Soulsbury & Legibe, 2018). It is not clear whether these inconsistencies actually reflect context-dependent HFCs (e.g. Arct et al., 2017; Ferrer, García-Nayas, Sanz, & Ortego, 2016), or rather stem from methodological limitations related to the sample size and/or the number of markers used (Colman & Slate, 2003), the study timespan (Velando, Barros, & Moran, 2015), the life-history stages considered (Bichet et al., 2019), or the type and number of traits assessed (Chapman et al., 2009).

Increasing the likelihood of detecting HFCs can be achieved by combining large sample sizes with mid- to long-term monitoring of fitness traits (e.g. Bichet et al., 2019; Velando et al., 2015). Unfortunately, this can be resource-demanding and logistically difficult, and thus, it is not surprising that comprehensive studies on HFCs are scarce. Moreover, there is still limited understanding of the balance of inbreeding and outbreeding depression in natural populations and the predominance of general versus local effects. It is worth noting that these two topics were recognized as important research avenues in this field almost a decade ago (Chapman et al., 2009; Szulkin et al., 2010; Szulkin & David, 2011).

Negative HFCs—a sign of outbreeding depression—are not commonly reported in natural populations (Judson et al., 2018), and although this may indeed reflect that outbreeding is less frequent than inbreeding, HFCs are recurrently assumed to be positive by most researchers (Szulkin & David, 2011). Testing for outbreeding depression should be standard practice in HFC studies (Marshall & Spalton, 2000), but this is seldom the case. As a result, there may be a publication bias towards positive HFCs. For instance, partitioning markers according to their position in functional or nonfunctional areas, or using MLH indexes that are more likely to capture outbreeding, are potentially useful methods that have been discussed and should be considered more often (Colman & Slate, 2003; Szulkin et al., 2010). In a similar way, local effects are found more regularly than general effects (Chapman et al., 2009), but rigorous tests for single-locus HFCs are not usually applied (but see, e.g. Bichet et al., 2019; Küpper et al., 2010; Olano-Marin et al., 2011a; Olano-Marin, Mueller, & Kempenaers, 2011b). This generates the question of whether the majority of local effects described so far are biologically meaningful or mere statistical artefacts (Szulkin et al., 2010). Replicate studies are needed to overcome the problems associated with multiple testing. Moreover, because local effects are hardly detected unless very strong (Szulkin et al., 2010), using a large marker panel or sampling markers close to selected genes are effective ways of increasing the chances of detection (e.g. Küpper et al., 2010). This, however, is not generally practiced in HFC studies.

The blue tit (Cyanistes caeruleus) is a well-studied species with regard to HFCs. Different populations and a variety of fitness-related
traits have been assessed, including juvenile and adult survival (Arct et al., 2017; Olano-Marín et al., 2011a), and numerous measurements of reproductive success (Foerster, Delhey, Johnsen, Lifjeld, & Kempenaers, 2003; García-Navas, Ortego, & Sanz, 2009; Olano-Marín et al., 2011b). Overall, these studies suggest that individual heterozygosity has a positive impact on fitness and that general effects are more frequent than local effects. To evaluate the generality of such patterns, however, it is necessary to either replicate or expand these studies. This is not only true because the strength and direction of some of the HFCs documented were age- (e.g. Olano-Marín et al., 2011a), sex- (e.g. Arct et al., 2017), or context-dependent (e.g. Ferrer et al., 2016), but because most of these studies genotyped birds at a limited set of loci (5–26; but see Olano-Marín et al., 2011a, 2011b). Furthermore, the number of scored markers known to be close to functional genomic areas was even lower, which may have prevented the detection of local effects (Szulkin & David, 2011).

In this study, we investigated HFCs in a population of blue tits that was intensively monitored during the winter and breeding seasons of 2014–2017, in southern Germany. This is presumed to be a panmictic and genetically diverse population, which is ideal for evaluating the potential presence of both inbreeding and outbreeding depression (Küpper et al., 2010). More importantly, we expanded the study of Olano-Marín et al. (2011b) on the correlation between heterozygosity and reproductive success in a blue tit population at Kolbeterberg, Vienna, Austria, following a slightly modified version of their protocol, and using the same marker panel and fitness-related traits, with some additions (see Materials and Methods). We genotyped 1,355 adult birds with a genome-wide set of 87 microsatellites. Using a comprehensive analysis based on multi- and single-locus heterozygosity measures, this study aimed to (a) evaluate the correlation between individual heterozygosity measures and six components of reproductive success, (b) determine whether HFCs in this population reflect the effects of inbreeding or outbreeding, and (c) test whether general or local effects underlie the observed results. We compare our findings with studies on other populations, particularly the one at Kolbeterberg (i.e. Olano-Marín et al., 2011b), and discuss them in regard to context-dependent HFCs.

2 | MATERIALS AND METHODS

2.1 | Study area and species

This research is part of a long-term study on the breeding biology of a population of blue tits carried out since 2007. The study population breeds in nest boxes mounted in a 40-ha mixed deciduous/co- niferous protected forest (“Westerholz,” 48°08′26″N, 10°53′29″E), within a larger (~600 ha) woodland in southern Germany. The study area has a higher density of oak trees (Quercus robur) compared to nearby forest patches and acts as a local “island” of high-quality breeding habitat for blue tits. The nearest forest with a similar high oak density is located ~20 km south-west of the study site. The study site contains 277 nest boxes, in which 60–170 blue tit pairs breed every year. Nest box occupation rate was 41%–64% during the study period, and breeding density was ~2–4 breeding pairs/ha. Natural cavities are available within the study site, but seem seldom used (in the Vienna population: 1.8% of the breeding population during 1998–2002; Foerster et al., 2003), presumably because nest boxes are preferred over tree cavities. The presence of nonbreeding floaters is hard to determine, but the occurrence of replacement polygyny after the disappearance of a breeding male suggests that their numbers during the breeding season might be low (Kempenaers, 1994).

The blue tit is a small passerine bird with a predominantly socially monogamous mating system, but with frequent extrapair paternity (Schlicht, Valcu, & Kempenaers, 2015). In our study population, blue tits produce one clutch per pair per breeding season, although replacement clutches after breeding failure are observed. Blue tits have a continuous distribution across Europe (Gosler & Clement, 2007). Populations in the northern part of the species range are partially migratory, whereas those in central Europe are considered to be resident (Stenning, 2018). Birds from nonmigratory populations are suspected to perform local movements in the nonbreeding season, leaving breeding areas to overwinter in nearby towns and villages (Dhondt, Adriaensen, & Plompen, 1995; Nilsson & Smith, 1988).

2.2 | Population monitoring and general procedures

Data for this study were collected during three “sampling seasons” between August 2014 and July 2017. A sampling season comprised a 12-month period between 1 August of a given year and 31 July of the following year. The breeding season in our study site spans nearly three months, from early March to mid-June.

All nest boxes were visited at least once per week during the breeding season. Data on laying date (date of first egg), clutch size and the number of hatchlings and fledglings were recorded for occupied nest boxes. Adults were caught in the nest box while feeding 8- to 10-day-old nestlings, marked with a uniquely numbered metal band and 1–3 coloured bands, aged (yearling or adult; Jenni & Winkler, 1994) and bled by brachial venipuncture. At day 14 after hatching, nestlings were also marked with a metal band and bled. A passive integrated transponder (PIT) was implanted subcutaneously on the back of all birds during first capture (Schlicht & Kempenaers, 2015). Each nest box was equipped with a transponder reader, two light barriers, a real-time clock and a data storage device (see Loës, Skrypski, & Kempenaers, 2019a), which thus allowed us to record visits from PIT-tagged birds throughout the year.

Each season between October and March we checked nest boxes at night at least once per month to catch unmarked birds that were roosting inside. We also deployed 16 feeders along the border of the study site and captured birds with mist nets. The feeders were equipped with a similar automated monitoring system as the nest boxes, enabling us to record PIT-tagged birds that used them during the winter (see Loës, Skrypski, & Kempenaers, 2019b).
We extracted DNA from blood samples for parentage analysis and molecular sexing. Parentage analysis was performed using 14 highly polymorphic microsatellite loci, following the methodology described in Appendix S1. As markers were highly informative (25 alleles per marker on average), the probability of erroneous assignment of males as sires was close to zero. The sex of all birds was assigned by amplifying the CHD locus with the primers P2/P8 (Griffiths, Double, Orr, & Dawson, 1998).

### 2.3 Components of reproductive success

Following Olano-Marín et al. (2011b), we included four proxies of individual reproductive success: (a) clutch size, as a measure of female fecundity; (b) male reproductive success, estimated as the number of offspring sired including both within- and extrapair young; (c) hatching; and (d) fledging success, calculated as the proportion of eggs in a clutch that hatched, and the proportion of nestlings that fledged, respectively.

We considered two additional variables related to reproductive success: (e) arrival date, defined as the first record of an individual in the breeding area during each season (see Gilsenan, Valcu, & Kempenaers, 2020); and (e) the probability of breeding locally, specified as a binary variable (breeder or nonbreeder), that is whether an individual ended up breeding in our study plot in a given season (see Bichet et al., 2019). Gilsenan et al. (2020) showed that arrival date at the breeding site is an individual trait that correlates negatively with reproductive success in this non-migratory blue tit population. Although we cannot exclude that some individuals classified as non breeders bred locally in natural cavities, the proportion of such cases in the breeding population is likely negligible (see above).

### 2.4 Microsatellite markers and loci categorization

We genotyped all captured yearlings and adults (n = 1,355; 670 females and 685 males) using a genome-wide set of 87 microsatellite markers developed for the blue tit (see details of primer sequences and PCR conditions in Olano-Marín et al., 2010). As these markers were originally mapped on the zebra finch and chicken genomes (see Olano-Marín et al., 2010), we here provide updated information on their characteristics, including chromosome location and position in relation to regions of gene expression/regulation in the blue tit genome (Mueller, Kuhl, Timmermann, & Kempenaers, 2016; Table S1). For this, we followed the methods outlined by Olano-Marín et al. (2010), performing a BLAT search of each microsatellite-containing sequence against the v2.0 blue tit genome, deposited at the Sequencing Core Facility of the Max Planck Institute for Molecular Genetics (http://public-genomes-NGS.molgen.mpg.de) and NCBI db (BioProject PRJNA284903).

All markers were categorized as “functional” and “nonfunctional” loci (F and NF; Olano-Marín et al., 2011a, 2011b) according to their location in the blue tit genome. Any locus contained within a transcribed sequence (RNAseq) was classified as F, with the exception of potentially spliced-out regions—that is introns. Nontranscribed loci, or intronic sequences within the blue tit transcriptome, were tagged as NF (Appendix S1).

We estimated the frequency of null alleles and tested for deviations from Hardy–Weinberg Equilibrium (HWE) at each locus, and tested for LD between all pairs of loci (see Appendix S1). For subsequent analyses, we used a final set of 83 loci (61 F and 22 NF loci; Table S1).

### 2.5 Population genetic structure

Assuming distinct genetic groups with varying levels of inbreeding/outbreeding as a panmictic population can increase—or obscure—the strength of HFCs (Chapman et al., 2009). We therefore used a principal components analysis (PCA) to explore and summarize genetic diversity in our study population, and to assess cryptic genetic substructure (Patterson, Price, & Reich, 2006). Additionally, we used the snapclust clustering algorithm (Beugin, Gayet, Pontier, Devillard, & Jombart, 2018) to estimate the number of genetic clusters present in the population. The snapclust rapidly converges to a maximum-likelihood solution by combining a geometric approach and the expectation-maximization (EM) algorithm, and performs comparably to standard clustering methods (Beugin et al., 2018).

We tested values of K—that is genetic clusters—between 1 and 10 and selected the optimal value using the Akaike information criterion (AIC; Akaike, 1998). We performed these analyses for the entire data set—that is all seasons combined—as well as for each sampling season, and each sex and age class (yearling or adult), using the adegenet package (Jombart, 2008) in the free software r 3.5.2 (R Core Team, 2018).

### 2.6 Heterozygosity measures

For analyses of general effects (see below), we used the homozygosity by locus (HL: Aparicio, Ortego, & Cordero, 2006) and the mean-square distance between alleles (md²; Coulson et al., 1998) indexes as measures of MLH. The HL index is commonly used in HFC studies as it shows a higher correlation with the inbreeding coefficient and with genome-wide heterozygosity compared to other measures (Aparicio et al., 2006). The md² index (reviewed in Hansson, 2010), on the other hand, appears to be a suitable complementary index to measure outbreeding (Coltman & Slate, 2003; Marshall & Spalton, 2000). Both indexes were calculated in R, implementing the GENHET function (Coulon, 2010) to calculate HL, and a custom routine to estimate md² based on the formula described in Coulson et al. (1998). Values of md² were log-transformed prior to statistical analyses (Coltman, Bowen, & Wright, 1998). MLH indexes were calculated with all markers, as well as with the F and NF loci separately.
For local effects, we calculated single-locus heterozygosity (SLH) by coding each locus as a binary variable, with “0” representing a homozygous state and “1” a heterozygous state.

2.7 | Estimating inbreeding and identity disequilibrium (ID)

We calculated the population mean value for the inbreeding coefficient $F$ by averaging the maximum-likelihood (ML) estimates of the inbreeding coefficient estimated for all genotyped individuals. ML values were computed using the adegenet package.

We used the $g_2$ statistic (David, Pujol, Viard, Castella, & Goudet, 2007) to assess ID—that is heterozygosity correlations across loci. Values of $g_2$ that statistically differ from zero indicate that heterozygosity is correlated with inbreeding and that HFCs are probably due to a genome-wide effect (Szulkin et al., 2010). We used the package inbreedR (Stoffel et al., 2016) to calculate $g_2$ and to test whether correlations across loci in our study population differed from the expectation under random association (see Appendix S1). We assessed ID for the entire population—that is breeders and nonbreeders—and the breeding population data sets, using all markers and also the F and NF loci separately (Olano-Marín et al., 2011a, 2011b).

2.8 | Testing HFCs: General effects

We assessed general effects by fitting mixed-effects models using the package lme4 (Bates, Maechler, Bolker, & Walker, 2015) in R. Because several individuals were recorded multiple times across seasons, we entered individual identity as a random intercept in all models (female id and/or male id, depending on the trait; see below). We included each measure of reproductive success as the response variable. As in Olano-Marín et al. (2011b), we fitted separate models for the F and NF loci. Results were considered statistically clear/conclusive whenever the 95% CI for the estimated effect sizes did not include zero (see Dushoff, Klein, & Bolker, 2019).

We used the breeding population data set to analyse the effect of heterozygosity on clutch size, male reproductive success and hatching/fledging success. For the first two traits, we fitted linear regressions after controlling for between-season effects by calculating z-scores using the mean value and standard deviation for each sampling season. For hatching/fledging success, we fitted binomial generalized linear models with clutch size and number of hatchlings, respectively, as the binomial denominator. All models included a measure of MLH and its squared term as predictors—HL and $md^2$ were modelled independently—and age as a nongenetic factor. Quadratic effects of MLH were excluded from all analyses as they showed a negligible effect on reproductive success. The models for hatching/fledging success included HL/$md^2$ and age of both parents as predictors, and female and male id as random effects. A potential interaction effect of female and male heterozygosity was further evaluated in a separate set of models (Table S2). We also assessed whether the effect of heterozygosity on hatching/fledging success changed when excluding nests with extrapair young from the analysis (Table S3). As these models yielded similar results to those presented below (see Results), they are not further discussed.

Following a similar approach, we analysed arrival date—for adult birds only to avoid bias in detection probability (see Appendix S1)—and the probability of breeding locally using linear regression and binomial models, respectively. The models for arrival date used the entire population and the breeding population data sets, while breeding probability was assessed only for the entire population. In both analyses, we accounted for whether a bird was a “local recruit”—that is a bird born in our study area—or an “immigrant” by adding a binomial predictor “status” (R or I). We also accounted for whether a bird had already bred in our study site by introducing a binomial predictor “previous breeding” (yes or no). Because early-arriving birds are more likely to occupy a nest box in our study site (Gilsenan et al., 2020), we entered “arrival period” (early or late) as covariate in the models of breeding probability. Following Gilsenan et al. (2020), we defined two arrival periods depending on whether a bird arrived before—early arrival—or after—late arrival—the median arrival date in a given season. We ran models separately for males and females. Details on model structure and sample sizes can be found in Appendix S1.

2.9 | Testing HFCs: Local effects

We tested whether the effect of heterozygosity on reproductive success was associated with specific loci. To this end, we ran similar models as described above, but using the SLH terms ($61$ for F loci models and $22$ for NF loci models) instead of the MLH measures. Some models had to be simplified due to convergence issues (Appendix S1).

We tested local effects by comparing the MLH and SLH models for each component of reproductive success, using likelihood ratio tests (Bichet et al., 2019; Küpper et al., 2010). Whenever we modified the SLH model to reach convergence, we applied the same modifications to the MLH model. We interpreted and further analysed local effects only when a SLH model explained variation in reproductive success better than the MLH model (Szulkin et al., 2010). To evaluate whether the observed local effects were consistent across measures of reproductive success—that is consistently positive or negative—we followed the meta-analytical approach of Coltman and Slate (2003) to calculate the mean effect size of each loci, using the Pearson product–moment correlation coefficient $r$ (Appendix S1).

3 | RESULTS

3.1 | General patterns of heterozygosity

The average expected heterozygosity ($H_e$) calculated from all markers was $0.57$ ($$\pm$$0.03), and the number of alleles per locus ($N_a$)
Measures of MLH showed lower genetic diversity for the F than for the NF loci (F loci: $H_L = 0.35, m^2 = 1.39$; NF loci: $H_L = 0.19, m^2 = 2.67$). $H_L$ and $m^2$ calculated based on all markers were weakly correlated ($r = -.35, p < .0001$), but measures obtained with only F or NF loci were independent ($r_{H_L} - m^2_F = 0.01, p = .81$; $r_{H_L} - m^2_{NF} = 0.03, p = .47$). Mean values of MLH did not differ between locally recruited birds and immigrants, but NF loci showed

**FIGURE 1** Genetic variation in a blue tit population in southern Germany. First three principal components from a PCA based on 1,355 multilocus genotypes. 80% of the total variation was explained by the first 175 dimensions. The first three components explained 1.54, 1.48 and 1.41% of the total variation, respectively. Genotyped individuals are represented according to (a) their demographic background (local recruits or immigrants) or (b) their breeding status (breeder or nonbreeder). Ellipses show 95% confidence intervals.
slightly higher diversity for birds that bred in the area at least once compared to those that did not (Figures S1 and S2).

### 3.2 | Population genetic structure

We did not find evidence for genetic substructure in the study population. The PCA based on the entire data set showed that all birds belonged to one genetic cluster (Figure 1). The same pattern was found when PCAs were performed separately for each sex and season (Figures S3 and S4). In the snapclust analysis of the entire population, the lowest AIC value was for $K = 1$, the highest for $K = 2$, with intermediate values for $K = 6$–$10$ (Figure 2). Similar results were obtained when the analysis was done for the breeding population only or for each sex and age class separately (Figures S5 and S6).

### 3.3 | Inbreeding and identity disequilibrium

The maximum-likelihood values of the individual inbreeding coefficient ranged from 0 to 0.23 ($F = 0$ for 52% of all genotyped individuals). The population mean value for the inbreeding coefficient was $F = 0.025$ (SD: ±0.04).

Heterozygosity was not correlated across loci in this population. The $g_2$ estimated for the entire population ($n = 1,355$) did not differ from zero (all loci: $g_2 = -0.00016$, 95% CI = $-0.0005$–$0.0002$, $p = .75$; F loci: $g_2 = -0.00025$, 95% CI = $-0.0011$–$0.0006$, $p = .70$; NF loci: $g_2 = -0.00059$, 95% CI = $-0.0014$–$0.0003$, $p = .89$). Moreover, the distribution of MLH in the population followed the expected pattern under random association of loci—that is when $g_2 = 0$ (Figure S7). The analyses of $g_2$ and MLH based on the breeding population only ($n = 601$) showed identical results (Figure S8).

#### FIGURE 2

Results from the snapclust analysis used to estimate the number of genetic clusters in a blue tit population in southern Germany. The most likely number of clusters (grey dot) was determined using Akaike’s information criterion for all recorded individuals ($n = 1,355$)

### 3.4 | Effects of MLH on reproductive success

Overall, analyses using HL and $m^2$ showed similar results, although the direction and size of the estimated effects did vary between F and NF loci (Figure 3). Regardless of the heterozygosity measure and the loci used, there was no statistical support for an effect of MLH on hatching success (Figure 3, Tables 1 and 2, Tables S4 and S5). For hatching success, the estimated effects varied depending on the type of loci used, thus providing inconclusive evidence (Figure 3). For the other measures of reproductive success, the effect of MLH was consistently negative, although some effects were not statistically conclusive (Figure 3).

For the F loci, we found negative HFCs for clutch size, arrival date and the probability of breeding locally (Figure 3, Table 1, Table S4). However, the estimated coefficients for female arrival date and male breeding probability were not statistically conclusive in all models (Figure 3). Interestingly, we found an interaction between MLH and arrival period on the probability of breeding: while the overall likelihood of occupying a nest box decreased as individual heterozygosity increased, heterozygous birds increased their chances of breeding in the study area when arriving later in the season (Table 1).

For the NF loci, the direction and size of the estimated effects of MLH showed high variability and were rather inconsistent (Figure 3, Table 2, Table S5). Results for arrival date and the probability of breeding locally followed the same pattern as in the F loci, but the 95% CI for these effects always included zero (Figure 3). Male reproductive success decreased with increasing heterozygosity, but this relationship was statistically consistent only when using HL as a measure of MLH (Figure 3).

### 3.5 | Effects of SLH on reproductive success

For F loci, log-likelihood ratio tests showed that SLH models fitted the data better for all reproductive traits except clutch size (Table 3). In contrast, for the NF loci, arrival date was the only variable in which single-locus effects were better supported. Regardless of the data set used—for example all birds or breeding birds only—AIC values gave higher support to SLH models with F loci than to those with NF loci (Table S6).

We tested single-locus HFCs in association with male reproductive success, arrival date and the probability of breeding locally. The probability of obtaining positive or negative single-locus effects was not different from 0.5 for any of the traits assessed (cumulative binomial tests; see also Figures S9–S11). The same was observed when considering only the statistically clear effects. Furthermore, we detected no difference in the proportion of negative HFCs between F and NF loci. Only locus "H095_TG05_053" (hereafter "H095") consistently showed a negative HFC across traits (Figures S9–S11).

The meta-analytic approach for calculating mean effect sizes for SLH included clutch size, male reproductive success, arrival date and the probability of breeding locally. Loci "H009_CcaTgu9" and "H089_CcaTgu29" (hereafter "H009" and "H089") showed a negative,
Using a genome-wide set of microsatellite markers and measures of reproductive success, we found evidence of negative, mostly local, HFCs in a blue tit population in southern Germany. Regardless of statistical congruence, HFCs based on MLH were consistently negative. Preliminary analyses showed that MLH was higher for birds that bred in the study area relative to nonbreeders, but differences were small and incongruent between F and NF loci. The frequency of negative versus positive single-locus effects was similar across markers, but the few loci that showed consistent local effects also showed a negative association with overall reproductive success. Clutch size was the only reproductive variable for which the general effect hypothesis was supported. For male reproductive success, arrival date and the probability of breeding, we found statistically clear, HFC in association with measures of female reproductive success (Figure 4). Locus “H095” showed a negative and relatively strong association with different proxies of reproductive success in males (Figure 5). When analysing both sexes combined, only locus “H095” showed a statistically clear negative HFC (Figure S12).

4 | DISCUSSION

Using a genome-wide set of microsatellite markers and measures of reproductive success, we found evidence of negative, mostly local,
evidence that certain loci contributed more heavily to the observed HFCs than others.

Overall, our study shows moderate correlations between reproductive success and measures of heterozygosity. In most HFC studies, the reported correlation coefficients ($r$) are typically small ($< 0.1$ as reviewed in Chapman et al., 2009; but see Velando et al., 2015). In this study, absolute values of $r$ ranged from 0.0002 to 0.25 for MLH—both HL and $md^2$ considered—and from 0.007 to 0.17 for single-locus effects.

4.1 | The importance of partitioning molecular markers

We found that, with few exceptions, only measures of MLH calculated from functional loci were associated with reproductive success. Similarly, consistent effects of SLH on female or male reproductive variables were only observed for functional loci. These results, along with the fact that we did not detect variation in inbreeding, indicate that most signals captured by MLH measures were due to local effects of a few functional loci (but see next section).

A study focused on effects of heterozygosity on reproductive success in a blue tit population at Kolbeterberg (Olano-Marín et al., 2011b). ~400 km to the east of our study site, reported some results that contrast with those reported here: (a) the Vienna population showed variation in inbreeding and strong evidence for general effects; and (b) the effects of MLH on fitness-related traits were only reflected in the nonfunctional loci (see Table 4). This difference illustrates how categorizing markers with presumably known characteristics can help understanding the genetic architecture of inbreeding and outbreeding depression (Szulkin & David, 2011). Putatively neutral markers are expected to reflect processes affecting genome-wide heterozygosity and thus might be better suited to detecting identity disequilibrium and general effects in populations with some degree of inbreeding, such as Kolbeterberg (Olano-Marín et al., 2011b). Functional markers, on the other hand, can facilitate the detection of relatively weak HFCs in genetically diverse populations due to their lower level of polymorphism (Küpper et al., 2010; Szulkin & David, 2011), such as Westerholz (this study). As these markers are expected to be closer to coding sequences, including polymorphisms responsible for outbreeding depression, they could capture the effects of outbreeding more efficiently than nonfunctional loci (Szulkin & David, 2011).

Our data suggest that blue tits in Westerholz comprise a panmictic, relatively outbred population with low variance in inbreeding. In this case, the observed negative effects of heterozygosity on reproductive success were reflected by functional loci, as expected. As stated by Olano-Marín et al. (2011b), the classification in functional and nonfunctional loci is not free of assumptions due to the lack of information regarding levels of expression of the microsatellite-containing sequences. Yet, the results from the two blue tit studies indicate that partitioning markers can provide further insight into the genetic mechanisms behind HFCs in animal populations. We adhere to Szulkin and David’s (2011) call encouraging researchers to follow this practice whenever possible.

4.2 | Negative general effects of heterozygosity

Our data support a general effect of heterozygosity on clutch size, even though we found no evidence for identity disequilibrium. Although theory predicts general effects to mostly occur in populations with variation in inbreeding, genome-wide HFCs can still be detected even when identity disequilibrium is not present (Kardos et al., 2014; Miller & Coltman, 2014). The main explanation for this is that fitness traits can potentially capture the effect of many more loci than those used in a given study, and as a result, weak effects of inbreeding—or outbreeding—can be detected through its phenotypic effects (Szulkin et al., 2010). This could explain the association between MLH and clutch size in this study. Our marker panel included microsatellites located on 23 different chromosomes, but 42% of these loci are on chromosomes 1, 2 and 7 (Table S1). It is thus likely that other loci that were not represented in our data set are responsible for these effects. Heterozygosity at the sampled loci should then be correlated with heterozygosity at genes affecting clutch size, but this remains to be verified.

Negative general effects are not common in the HFC literature (but see Bichet et al., 2019; Küpper et al., 2010), probably because they are not as ubiquitous as positive HFCs in natural populations. Alternatively, negative HFCs may be underreported (see Chapman et al., 2009; Hedrick, 2012), which could be related to the difficulty of interpreting outbreeding depression (Szulkin & David, 2011). Negative HFCs are frequently associated with single-locus effects, but explanations for genome-wide negative effects also exist. Multilocus homozygotes can have higher fitness compared to heterozygotes as a consequence of the association between genetic markers and fitness loci, a process known as associative underdominance (AU; Zouros, 1993). The effect of AU depends on the frequency and the degree of recessiveness of the alleles involved and the population inbreeding level (Deng & Fu, 1998). Under a noninbreeding scenario, the likelihood of detecting negative HFCs depends on the relative abundance of partial recessive deleterious mutations (see Mueller et al., 2011). It has been suggested that the majority of mutations with small fitness effects are partially recessive and could therefore contribute to negative HFCs (Eyre-Walker & Keightley, 2007; Mueller et al., 2011).

Negative general effects of heterozygosity have rarely been reported in blue tits (see Ferrer et al., 2016). Olano-Marín et al. (2011a) reported a negative effect of MLH on the probability of hatching and local recruitment of young females. Nevertheless, previous studies focused on adult reproductive success detected identity disequilibrium and/or showed positive HFCs (Ferrer et al., 2016; Foerster et al., 2003; Foerster, Valcu, Johnsen, & Kempenaers, 2006; García-Navas et al., 2009; Olano-Marín et al., 2011b). These differences in the direction of HFCs can be explained by historical factors, as populations have different
TABLE 1  Effects of heterozygosity (estimated as HL) and age (yearling or adult) on measures of reproductive success in a blue tit population in southern Germany. Results are from analyses based on 61 functional loci used to calculate HL. Linear mixed-effects models were fitted for clutch size, male reproductive success and arrival date. Binomial models were used for hatching success, fledging success and the probability of breeding locally. All models included individual identity as a random effect.

| Dependent variable | Clutch size | Male reproductive success | Hatching success | Fledging success |
|--------------------|-------------|---------------------------|-----------------|-----------------|
| Sample             | Breeding females (n = 318) | Breeding males (n = 343) | Breeding pairs (n = 278) | Breeding pairs (n = 230) |
| Parameter          | Estimate ± SE | t | p-Value | Estimate ± SE | t | p-Value | Estimate ± SE | Z | p-Value | Estimate ± SE | Z | p-Value |
| Intercept          | -0.87 ± 0.31 | -2.78 | .010 | -1.09 ± 0.56 | -1.93 | .060 | 6.91 ± 0.64 | 10.78 | .390 | 6.43 ± 1.63 | 3.96 | .310 |
| HL                 | 0.27 ± 0.12 | 2.38 | .018 | -0.003 ± 0.19 | -0.02 | 990 | -0.05 ± 0.24 | -0.19 | 980 | -0.43 ± 0.48 | -0.91 | .360 |
| Age (adult)        | 0.60 ± 0.21 | 2.87 | .005 | 0.76 ± 0.37 | 2.04 | .042 | -0.39 ± 0.22 | -1.75 | .080 | 0.12 ± 0.48 | 0.25 | .800 |
| HLFEMALES          | -0.05 ± 0.24 | -0.19 | .850 | -0.43 ± 0.48 | -0.91 | .360 | 0.62 ± 0.88 | 0.70 | .480 |
| HLMALES            | -2.18 ± 0.37 | -5.94 | <.001 | 0.62 ± 0.88 | 0.70 | .480 |
| AgeFEMALES (adult) | -1.83 ± 0.35 | -5.21 | <.001 | 0.77 ± 0.86 | 0.89 | .370 |
| AgeMALES (adult)   | -1.83 ± 0.35 | -5.21 | <.001 | 0.77 ± 0.86 | 0.89 | .370 |
| Random variance | 42.9% | 5.2% | 51.1%/ 20.8% | 60.9%/ 24.9% |

| Dependent variable | Arrival date | Probability of breeding locally |
|--------------------|--------------|---------------------------------|
| Sample             | All birds (n = 635) | All females (n = 294) | All males (n = 341) | Breeding females (n = 151) | Breeding males (n = 159) |
| Parameter          | Estimate ± SE | Z | p-Value | Estimate ± SE | Z | p-Value | Estimate ± SE | Z | p-Value | Estimate ± SE | Z | p-Value |
| Intercept          | -11.56 ± 5.29 | -2.19 | .030 | -9.39 ± 8.01 | -1.17 | .240 | -17.64 ± 6.44 | -2.74 | .010 | -28.39 ± 17.01 | -1.67 | .100 |
| HL                 | -2.72 ± 2.97 | -0.92 | .360 | -3.33 ± 4.56 | -0.73 | .470 | -5.27 ± 3.50 | 1.50 | .130 | -16.8 ± 6.76 | -2.49 | .014 |
| Local recruit (R)  | -5.33 ± 6.02 | -0.89 | .380 | 6.91 ± 6.65 | 0.72 | .480 | 0.21 ± 7.08 | 0.03 | .980 | 4.04 ± 13.79 | 0.29 | .770 |
| Prev.breed (yes)   | -48.30 ± 5.44 | -8.87 | <.001 | -30.23 ± 8.49 | -3.56 | <.001 | -68.43 ± 6.58 | -10.40 | <.001 | -9.61 ± 16.53 | -0.58 | .560 |
| Random variance & 63.6% | 61% | 57.7% | 73.9% | 78.7% | (Continues)
histories of inbreeding and demography, and hence distinct genetic architectures (Coltman & Slate, 2003). In blue tit populations in Spain (e.g. Garcia-Navas et al., 2009) and Austria (e.g. Foerster et al., 2003; Olano-Marín et al., 2011b), matings between close relatives have been documented and can generate a genetic substructure (see Table 4). Conversely, we did not find evidence for genetic structure that would suggest the occurrence of nonrandom mating, not even at a micro-geographic scale (i.e. no fine-scale genetic structure; Appendix S3-1). Under these circumstances, the occurrence of positive HFCs is unlikely, because inbreeding is probably negligible.

Differences in ecological factors between populations may also be responsible for the contrasting HFCs. A reasonable ecological explanation for negative effects of MLH on reproductive success is that local, less outbred individuals are more efficient at exploiting resources from their native habitats as they are better adapted than immigrants (Dias & Blondel, 1996; García-Navas, Ferrer, Sanz, & Ortego, 2014; Postma & van Noordwijk, 2005). For instance, more homozygous female black grouse (Lyrurus tetrix) laid larger clutches and had greater body mass than their more heterozygous conspecifics, which may be a consequence of the former having more access to resources critical for reproduction (Soulsbury & Legibre, 2018). In blue tits, Ferrer et al. (2016) detected transient negative HFCs in highly productive years and suggested that local, more homozygous individuals might outcompete more heterozygous individuals under relatively benign environmental conditions. Such context-dependent variation in HFCs may be frequent in fluctuating environments (Mueller et al., 2011), with homozygotes being more successful when selection on heterozygosity is temporarily relaxed (Armbruster & Reed, 2005; Fox & Reed, 2011). Breeding data from blue tits at Westerholz from 2007 to 2017 show that measures of breeding productivity—that is number of breeding pairs and hatching and fledging success—varied between years, and that our study coincided with a period of high breeding density (Appendix S3-2). This would support the idea of an "episodic homozygote advantage," but this needs to be examined in more detail.

The homozygote advantage in our Westerholz population was independent of whether an individual was a local recruit or an immigrant—defined here as an individual that did not fledge from a nest box. Heterozygosity did not differ between immigrants and recruits, and the results from the MLH model for clutch size did not change when including this categorization as a predictor (Appendix S3-3). This would suggest that any existing local—homozygote—advantage in this population is not related to an individual’s immigration status, perhaps because many “immigrants” are birds that were raised in neighbouring forests.

It could be argued that the contrasting HFCs observed among blue tit populations are due to methodological differences between studies. Although it is difficult to determine whether that is the case, it is possible to test for between-population differences by using a standardized analysis. We combined our data set and the multi-locus genotypes used by Olano-Marín et al. (2011b) to assess the effect of MLH on clutch size in Westerholz and Kolbeterberg. The
TABLE 2 Effects of heterozygosity (estimated as HL) and age (yearling or adult) on measures of reproductive success in a blue tit population in southern Germany. Results are from analyses based on 22 nonfunctional loci used to calculate HL. Linear mixed-effects models were fitted for clutch size, male reproductive success and arrival date. Binomial models were used for hatching success, fledging success and the probability of breeding locally. All models included individual identity as a random effect.

| Dependent variable                  | Clutch size                           | Male reproductive success | Hatching success                                | Fledging success                          |
|-------------------------------------|---------------------------------------|---------------------------|------------------------------------------------|-------------------------------------------|
| Sample                              | Breeding females (n = 318)            | Breeding males (n = 343)  | Breeding pairs (n = 278)                         | Breeding pairs (n = 230)                 |
| Parameter                           | Estimate ± SE                         | Estimate ± SE             | Estimate ± SE                                     | Estimate ± SE                             |
|                                     | t                                     | t                         | Z                                               | Z                                          |
|                                     | p-Value                               | p-Value                   | p-Value                                         | p-Value                                   |
| Intercept                           | −0.86 ± 0.31                          | −1.05 ± 0.56              | 6.84 ± 0.63                                      | 6.60 ± 1.59                               |
|                                     | −2.78                                 | −1.87                     | 10.86                                           | 4.15                                      |
| HL                                  | 0.18 ± 0.12                           | 0.38 ± 0.19               | 0.17 ± 0.22                                      | 0.18 ± 0.17                              |
|                                     | 1.52                                  | 2.06                      | 0.76                                            | 1.02                                     |
|                                     | .13                                   | .041                      | .45                                             | .31                                       |
| Age (adult)                         | 0.59 ± 0.21                           | 0.73 ± 0.37               | −0.008 ± 0.24                                    | −0.25 ± 0.44                             |
|                                     | 2.81                                  | 1.99                      | −0.03                                           | −0.58                                    |
|                                     | .005                                  | .48                       | .97                                             | .56                                       |
| HLFEMALES                           | 0.58 ± 0.22                           | 0.73 ± 0.37               | −6.29 ± 4.59                                     | 0.59 ± 0.88                              |
|                                     | 2.64                                  | 0.64                      | −1.37                                           | 0.67                                      |
|                                     | .008                                  | .39                       | .76                                             | .51                                       |
| AgeFEMALES (adult)                  | −2.16 ± 0.36                          | −5.24                      | 5.79 ± 14.19                                     | 0.41                                      |
|                                     | −6.03                                 | −5.24                      | −3.67 ± 7.04                                     | .68                                       |
|                                     | <.001                                 | <.001                      | −0.52                                           | .80                                       |
|                                     |                                      |                           | −0.86                                           | .39                                       |
| AgeMALES (adult)                    | −1.81 ± 0.34                          | −5.24                      | −3.67 ± 7.04                                     | −0.86                                    |
|                                     | −6.03                                 | −5.24                      | −0.52                                           | −0.38                                    |
|                                     | <.001                                 | <.001                      | .67                                             | .86                                       |
|                                     |                                      |                           | .51                                             | .39                                       |
| Random variance^a                   | 44%                                   | 3.1%                      | 53.9%                                           | 60.9%                                    |
|                                     |                                       |                           | 17.7%                                           | 23.1%                                    |

| Dependent variable                  | Arrival date                          | Probability of breeding locally |
|-------------------------------------|---------------------------------------|---------------------------------|
| Sample                              | All birds (n = 635)                   | All females (n = 294)           | All males (n = 341)                            |
| Parameter                           | Estimate ± SE                         | Estimate ± SE                   | Estimate ± SE                                  |
|                                     | t                                     | t                               | Z                                               |
|                                     | p-Value                               | p-Value                         | p-Value                                         |
| Intercept                           | −11.85 ± 5.29                         | −9.13 ± 7.99                    | −18.33 ± 6.46                                  |
|                                     | −2.24                                 | −1.14                           | −2.84                                           |
|                                     | .43                                   | .17                            | .77                                             |
| Local recruit (yes)                 | −5.53 ± 6.03                          | 6.18 ± 9.64                     | −0.08 ± 7.13                                    |
|                                     | −0.92                                 | 0.64                           | −0.01                                           |
|                                     | .36                                   | .52                            | .99                                             |
| Prev. breed (yes)                   | −47.79 ± 5.43                         | −30.59 ± 8.47                   | −67.02 ± 6.55                                   |
|                                     | −8.80                                 | −3.61                          | −10.23                                          |
|                                     | <.001                                 | <.001                          | <.001                                           |
|                                     |                                      |                                 | −10.59 ± 16.79                                  |
|                                     |                                      |                                 | −0.63                                           |
|                                     |                                      |                                 | .53                                             |
|                                     |                                      |                                 | −24.86 ± 12.67                                  |
|                                     |                                      |                                 | −1.96                                           |
|                                     |                                      |                                 | .053                                            |
| Random variance^a                   | 63.8%                                 | 61.4%                          | 58.4%                                           |
|                                     |                                       |                                 | 75.7%                                           |
|                                     |                                       |                                 | 79.7%                                           |

(Continues)
population-dependent sign of HFCs—negative in Westerholz and positive in Kolbeterberg—was further supported by this analysis (Figure 6; see also Appendix S3-4).

### 4.3 Negative local effects of heterozygosity

We found support for negative, local effects of heterozygosity on arrival date, the probability of breeding locally and male reproductive success. Thus, the signal captured by MLH for these traits was most likely due to the contribution of a few loci. Our results support the idea that most HFCs observed in noninbred populations might be based on local effects at single loci instead of multilocus effects (see Chapman et al., 2009; Hansson & Westerberg, 2008; Mueller et al., 2011).

Single-locus effects are always expected given the high error variance in their estimation, and they must be interpreted carefully (Szulkin et al., 2010). The best strategy to find biologically meaningful effects is to examine the distribution of local effect sizes (Chapman et al., 2009; Szulkin et al., 2010), to identify atypically large effects (see, e.g. García-Navas, Cáliz-Campal, Ferrer, Sanz, & Ortego, 2014) and to check for consistent effects in terms of direction and/or size. In our study, the distribution of effect sizes did not suggest a prevalence of either positive or negative values for any of the reproductive variables assessed. Because different measures of reproductive success are likely correlated with some degree, we expected biologically important effects to be consistently negative or positive. Although some loci showed comparatively strong effects, the direction and/or size of these were incongruent in most cases. Although loci showed comparatively strong effects, the direction and/or size of these were incongruent in most cases. Only locus "H095" showed a consistent and relatively strong, negative effect. This was true for the regression analyses fitted independently for each reproductive trait, as well as for the meta-analytical approach used to calculate the correlation coefficient \( r \). The effect was observed when analysing all individuals or males only, suggesting that variation in this locus might be relevant for male blue tits. The meta-analysis further suggested that loci "H009" and "H089" were important for females, but they did not show any effect in the regression models fitted for each variable.

Although Olano-Marín et al. (2011b) found strong evidence for general effects in the Austrian blue tit population, they also reported potential local effects for these three loci (Table 4). In their study, "H009" also had a negative effect on female reproductive success, specifically on fledging success. Conversely, "H089" showed a positive effect on hatching success, but only for males. Remarkably, "H095" negatively affected siring success and it was the second strongest local effect. Thus, data from populations in Germany and Austria point to locus "H095" as an interesting marker associated with a homozygote advantage for male reproductive success.

Local effects are theoretically attributed to linkage disequilibrium between the microsatellite markers responsible for HFCs and fitness loci (Szulkin et al., 2010). This could also be the case for locus "H095," but we cannot rule out the possibility of physical linkage with a coding gene. The microsatellite-containing sequence...
### Table 3
Comparisons between single-locus (SLH) and multilocus (MLH) models for testing effects of heterozygosity on measures of reproductive success in a blue tit population in southern Germany. Results from log-likelihood ratio tests are reported for models based on functional and nonfunctional loci.

| Functional loci (61) | Clutch size (CS) | Nonfunctional loci (22) | Clutch size (CS) |
|----------------------|------------------|-------------------------|------------------|
| **Sample**           |                  |                         |                  |
| Breeding females (n = 318) |                  | Breeding females (n = 318) |                  |
| **Model**            | df | LogLik | test df | χ² | p-value | df | LogLik | test df | χ² | p-value |
| CS – MLH + age + (1 | 5 | -662 |          |    | .69     | 5 | -664 |          |    | .50     |
| CS – SLH + age + (1 | 65 | -635 | 60  | 54.18 | .001 | 66 | -654 | 21  | 20.27 | .001 |
| Male reproductive success (MRS) |                  |                          |                  |
| Breeding males (n = 343) |                  | Breeding males (n = 343) |                  |
| MRS – MLH + age + (1 | 5 | -908 |          |    | .001 | 5 | -906 |          |    | .001 |
| MRS – SLH + age + (1 | 65 | -856 | 60  | 104.27 | <.001 | 66 | -890 | 21  | 31.49 | .066 |
| Arrival date (AD)   |                  |                          |                  |
| All birds (n = 635) |                  |                          |                  |
| AD – MLH + R + PrevB + (1 | 6 | -3507 |          |    | .001 | 6 | -3429 | 21  | 156.07 | <.001 |
| AD – SLH + R + PrevB + (1 | 66 | -3289 | 60  | 436.73 | <.001 | 27 | -3429 | 21  | 156.07 | <.001 |
| Breeding birds (n = 310) |                  |                          |                  |
| AD – MLH + R + PrevB + (1 | 6 | -1700 |          |    | .001 | 6 | -1694 | 21  | 177.29 | <.001 |
| AD – SLH + R + PrevB + (1 | 66 | -1449 | 60  | 502.49 | <.001 | 27 | -1614 | 21  | 177.92 | <.001 |
| All females (n = 294) |                  |                          |                  |
| AD – MLH + R + PrevB + (1 | 6 | -1630 |          |    | .001 | 6 | -1629 |          |    | .001 |
| AD – SLH + R + PrevB + (1 | 66 | -1392 | 60  | 477.43 | <.001 | 27 | -1543 | 21  | 172.44 | <.001 |
| All males (n = 341) |                  |                          |                  |
| AD – MLH + R + PrevB + (1 | 6 | -1835 |          |    | .001 | 6 | -1836 |          |    | .001 |
| AD – SLH + R + PrevB + (1 | 66 | -1605 | 60  | 460.96 | <.001 | 27 | -1757 | 21  | 157.56 | <.001 |
| Breeding females (n = 151) |                  |                          |                  |
| AD – MLH + R + PrevB + (1 | 6 | -831 |          |    | .001 | 6 | -834 |          |    | .001 |
| AD – SLH + R + PrevB + (1 | 66 | -561 | 60  | 541.10 | <.001 | 27 | -739 | 21  | 188.6  | <.001 |
| Breeding males (n = 159) |                  |                          |                  |
| AD – MLH + R + PrevB + (1 | 6 | -832 |          |    | .001 | 6 | -835 |          |    | .001 |
| AD – SLH + R + PrevB + (1 | 66 | -572 | 60  | 518.79 | <.001 | 27 | -746 | 21  | 176.03 | <.001 |

(Continues)
of “H095” is located within the exon of the MDGA2 protein coding gene (Table S1). The expression of this gene is biased towards the brain and the testes in different vertebrates, and is apparently associated with phenotypic differences in body mass and cognitive function in humans (Stelzer et al., 2016). Further study is needed to determine whether a polymorphism in locus “H095” is a good predictor of transcriptional variation in MDGA2 in this blue tit population.

Negative HFCs can arise from linkage disequilibrium between molecular markers and functional loci displaying (a) underdominance, (b) codominant allele advantage, and (c) recessive allele advantage (Kardos et al., 2014; Mueller et al., 2011). Negative HFCs caused by underdominance are only stable if the less frequent genotype at the locus of interest is favoured (Mueller et al., 2011), but we did not observe any reproductive advantage for individuals carrying the less common genotype at locus “H095” (Appendix S3-5). Recessive allele advantage may also be the cause, but this is unstable in natural populations and recessive alleles are rarely picked up by selection and reach fixation (Mueller et al., 2011). Negative HFCs can also arise at codominant loci when the beneficial allele(s) is common (Mueller et al., 2011). Note that the two most frequent alleles (allele frequency: 204 = 0.36; 208 = 0.29) in their homozygote forms comprised ~90% of all homozygotes at locus “H095” and that ~ 50% of all breeding males were either 204/204 or 208/208 (Appendix S3-5). These two alleles were also the most common variants at locus “H095” in the blue tit population in Austria (details not shown).

### 4.4 A genetic effect on timing of arrival?

The negative association between heterozygosity at locus “H095” and male arrival date, probability of breeding and reproductive success can be interpreted in two ways: a pleiotropic effect of the locus—or loci—linked to “H095,” or an effect on one variable that is correlated with the other two. Although we cannot dismiss the first explanation, the second one may be more likely. Firstly, heterozygosity at locus “H095” showed the largest effect size for arrival date. Secondly, males that arrived earlier in our study site had higher reproductive success. For instance, early-arriving males were more likely to breed locally (see Gilsenan et al., 2020). This may be related to competition for procuring a nest box in this high-quality breeding habitat (see Study area and species), as only 32%–44% of all males that were detected during the nonbreeding season were later identified as breeders. In fact, ~62% of late-arriving individuals did not obtain a territory and were likely forced to breed elsewhere—or died. Among breeding males, individuals that arrived earlier were also more likely to sire extrapair young (Gilsenan et al., 2020). Together, these results imply that a potential genetic effect on the timing of arrival could have repercussions for future reproductive success—that is carry-over effects. Carry-over effects of timing of arrival are apparently important for migratory birds (Tarka, Hansson, & Hasselquist, 2015; Tryjanowski, Sparks, Ptaszyn, & Kosicki, 2004; see reviews by Alerstam, 2011; Kokko, 1999), and this might also be the case for nonmigratory species (see Gilsenan...
In our study population, timing of arrival is an individual-specific, repeatable trait that is related to mating and that potentially reflects individual quality (Gilsenan et al., 2020). Therefore, selection on this behaviour will likely affect other traits related to reproduction.

Timing of arrival has only been considered in a few HFC studies, particularly those involving migratory species (e.g. Bichet et al., 2019). This is reasonable, because variation in migratory behaviour is partially explained by genetic effects (Pulido, 2007). It is possible, however, that timing of arrival of individuals from resident populations is also under genetic control. For example, the movement behaviour could simply be expressed on a smaller geographical scale (Gilsenan et al., 2020), such that variation in timing of arrival reflects the distance between the breeding and nonbreeding sites. Alternatively, but not mutually exclusive, individuals might differ intrinsically in the timing of their return to the breeding site, even if they overwinter in the same area—or at an equal distance from the breeding site. We speculate that the polymorphism at locus "H095" is either directly related to phenological traits or
that it reflects individual differences in the distance between overwintering areas and breeding sites (see Judson et al., 2018; Knapp, Prince, & James, 2016). It would be useful to assess the role of MDGA2 and other genes linked to "H095" in the expression of differences in timing of arrival.

Because arrival date also affects female reproductive success in our study population—for example clutch size (Gilsenan et al., 2020), and we also observed large effects of MLH on female arrival date, we cannot exclude the possibility that the general effects captured in this study are also a consequence of the timing of arrival to the breeding area.

5 | CONCLUDING REMARKS

This study shows evidence for negative, mostly local effects of heterozygosity on reproductive success in a genetically homogeneous, panmictic population of blue tits. Negative HFCs are probably more
**TABLE 4** Comparison of genetic population structure and effects of multilocus (MLH) and single-locus (SLH) heterozygosity on reproductive success between two blue tit populations. The effects of MLH (estimated as HL) are summarized for variables of reproductive success measured at both sites. MLH was calculated using either 61/58 functional (F) or 22/21 nonfunctional (NF) loci. Nonconclusive (NC) evidence corresponds to cases where the estimated effect sizes did not differ significantly from zero (overlapping 95% CI). The effects of SLH are summarized only for loci that showed an effect on reproductive success in this study (see main text).

| Within-population genetic structure | Population          |
|-------------------------------------|---------------------|
|                                     | Westerholz (Germany) | Kolbeterberg (Austria) |
| Identity disequilibrium (ID)        | Not detected        | Detected               |
| Evidence for single-locus effects   | Strong              | Weak                   |

**Multilocus effects**

| Trait                        | F loci: Negative/ NF loci: NC | F loci: NC/ NF loci: Positive |
|------------------------------|-------------------------------|-------------------------------|
| Clutch size                  | F loci: NC/ NF loci: Negative |
| Male reproductive success    | F loci: NC/ NF loci: Negative |
| Hatching success             | F loci: NC/ NF loci: Negative |
| Fledging success             | F loci: NC/ NF loci: NC       |

**Single-locus effects**

| Locus                        | Negative (females) | Positive (males) |
|------------------------------|--------------------|------------------|
| H009_CcaTgu9                 |                    |                  |
| H089_CcaTgu29                | Negative (females) |                  |
| H095_TG05_053d               | Negative (males)   |                  |

\(^a\)This study.

\(^b\)Olano-Marín et al. (2011b).

\(^c\)Only clutch size was formally compared between the two populations as it was the only trait for which we did not find evidence for a single-locus effect.

\(^d\)Note that this is the only locus for which we found a consistent, negative effect across measures of reproductive success.

**FIGURE 6** Relationship between individual heterozygosity at multiple loci (MLH) and female clutch size (z-scores, based on the mean and standard deviation for each sampling season) in two populations of blue tits from Germany (n = 318; this study) and Austria (n = 408; Olano-Marín et al., 2011b). MLH was measured as HL based on 73 loci that were scored at both populations. Shown are the raw data (points) and predictions with 95% confidence intervals from a linear mixed-effects model (see Appendix S3-4 for more details). The direction of the effect of MLH on clutch size is dependent on the population (HL*Population: estimate −0.74, 0.19 SE; t = −3.84, p < .001).
frequent in natural populations than the literature suggests (Judson et al., 2018; Marshall & Spalton, 2000), as could local adaptation and outbreeding depression (Verhoeven et al., 2010). Alternatively, recurrent mutations with specific dominant/recessive fitness effects could be responsible for negative HFCs in noninbreeding scenarios (Mueller et al., 2011). More studies in relatively outbred populations are warranted (Chapman et al., 2009) to test whether negative HFCs are more common (Küpper et al., 2010). Our results suggest that characterizing and partitioning markers into functional categories is a valuable practice that can help understand the genetic architecture of inbreeding and outbreeding depression (Szulkin & David, 2011).

We suggest that negative HFCs in this blue tit population are likely due to a local genetic effect on arrival date at the breeding site. Further study is required to test whether this is a local advantage of less outbred individuals (see Marr, Keller, & Arcese, 2002; Marshall & Spalton, 2000), or a temporary pattern only detectable under favourable environmental conditions (Ferrer et al., 2016). Our results and those from Olano-Marín et al. (2011b) reveal that the occurrence of positive and negative HFCs varies intraspecifically, possibly as a result of population differences in historical demographic and ecological factors. Comparative approaches are essential to discern how spatial and temporal variation in HFCs results from a balance of inbreeding and outbreeding depression. We suggest future studies should consider different life-history stages (as in Bichet et al., 2019) and assess long-term fitness benefits in relation to heterozygosity. This would be a step forward in understanding the coexistence of positive and negative HFCs in populations and the selective forces driving heterozygote and homozygote advantage.

ACKNOWLEDGEMENTS

We thank all members of the field team, and in particular Agnes Türk and Andrea Wittenzellner, for catching blue tits and monitoring their breeding behaviour at Westerholz, Sylvia Kuhn for genotyping, Juanita Olano-Marín for providing data that facilitated the interpretation of the results from the blue tit population at Kolbeterberg, and David Coltman, Vicente García-Navas and two anonymous reviewers for constructive comments on the manuscript. Funding for this study was provided by the Max Planck Society to B.K. Open access funding enabled and organized by ProjektDEAL.

AUTHOR CONTRIBUTIONS

The study was conceived by B.K. The analytical methodology was devised and the data were analysed by E.B.-D. E.B.-D. wrote the manuscript with input from B.K. and with edits from J.C.M. and C.G. Field data were collected by C.G.

DATA AVAILABILITY STATEMENT

Detailed characteristics of microsatellite loci can be found in Supplemental Information Table S1. Files with raw genotypes, SLH and MLH genotypes, and measures of reproductive success are available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.4f4qrjf8t).

ORCID

Esteban Botero-Delgadillo https://orcid.org/0000-0003-4653-7551
Carol Gilsenan https://orcid.org/0000-0002-6054-1461
Jakob C. Mueller https://orcid.org/0000-0001-6676-7595
Bart Kempenaers https://orcid.org/0000-0002-7505-5458

REFERENCES

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), Selected papers of Hirotugu Akaike (pp. 199–213). Springer Series in Statistics, New York, NY: Springer.
Alerstam, T. (2011). Optimal bird migration revisited. Journal of Ornithology, 152(supp 1), 55–223. https://doi.org/10.1007/s1033 6-011-0694-1
Aparicio, J. M., Ortego, J., & Cordero, P. J. (2006). What should we weigh to estimate heterozygosity, alleles or loci? Molecular Ecology, 15(14), 4659–4665. https://doi.org/10.1111/j.1365-294X.2006.03111.x
Arct, A., Sudyka, J., Podmoka, E., Drobniai, S. M., Gustafsson, L., & Cichon, M. (2017). Heterozygosity-fitness correlations in blue tit nestlings (Cyanistes caeruleus) under contrast in rearings conditions. Evolutionary Ecology, 31(5), 803–814. https://doi.org/10.1007/s1068 2-017-9911-6
Armbruster, P., & Reed, D. H. (2005). Inbreeding depression in benign and stressful environments. Heredity, 95, 235–242. https://doi. org/10.1038/sj.hdy.6800721
Balloux, F., Amos, W., & Coulson, T. (2004). Does heterozygosity estimate inbreeding in real populations? Molecular Ecology, 13(10), 3021–3031. https://doi.org/10.1111/j.1365-294X.2004.02318.x
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01
Beugin, M.-P., Gayet, T., Pontier, D., Devillar, S., & Jombart, T. (2018). A fast likelihood solution to the genetic clustering problem. Methods in Ecology and Evolution, 9(4), 1006–1016. https://doi. org/10.1111/2041-210X.12968
Bichet, C., Vedder, O., Sauer-Gürtl, H., Becker, P. H., Wink, M., & Bouwhuis, S. (2019). Contrasting heterozygosity-fitness correlations across life in a long-lived seabird. Molecular Ecology, 28(3), 671–685. https://doi.org/10.1111/mec.14979
Boerner, M., Hoffman, J. I., Amos, W., Chakarov, N., & Kruger, O. (2013). No correlation between multi-locus heterozygosity and fitness in the common buzzard despite heterozygote advantage for plumage co- lour. Journal of Evolutionary Biology, 26(10), 2233–2243. https://doi. org/10.1111/jeb.12221
Chapman, J. R., Nakagawa, S., Coltman, D. W., Slate, J., & Sheldon, B. C. (2009). A quantitative review of heterozygosity-fitness correlations in animal populations. Molecular Ecology, 18(13), 2746–2765. https://doi.org/10.1111/j.1365-294X.2009.04247.x
Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. Nature Reviews Genetics, 10, 783–796. https://doi. org/10.1038/nrg2664
Coltman, D. W., Bowen, W. D., & Wright, J. M. (1998). Birth weight and neonatal survival of harbor seal pups are positively correlated with genetic variation measured by microsatellites. Proceedings of the Royal Society of London B-Biological Sciences, 265(1398), 803–809. https://doi.org/10.1098/rspb.1998.0363
Coltman, D. W., & Slate, J. (2003). Microsatellite measures of in-breeding: A meta-analysis. Evolution, 57(5), 971–983. https://doi.org/10.1111/j.0004-3802.2003.tb00309.x
Coulon, A. (2010). GENHET: An easy-to-use R function to estimate in- dividual heterozygosity. Molecular Ecology Resources, 10(1), 167–169. https://doi.org/10.1111/j.1755-0998.2009.02731.x
Loës, P., Skrypski, P., & Kempenaers, B. (2019b). GitHub repository MOMO, link: https://github.com/peterloes/MOMO. https://doi.org/10.5281/zenodo.3516588

Lynch, M. (1991). The genetic interpretation of inbreeding depression and outbreeding depression. Evolution, 45(3), 622–629. https://doi.org/10.1111/j.1558-5646.1991.tb04333.x

Lynch, M., & Walsh, B. (1998). Inbreeding depression in reintroduced Arabian oryx. Animal Conservation, 3(3), 241–248. https://doi.org/10.1111/j.1469-1795.2000.tb00109.x

Miller, J. M., & Coltman, D. W. (2014). Assessment of identity disequilibrium and its relation to empirical heterozygosity-fitness correlations: A meta-analysis. Molecular Ecology, 23(8), 1899–1909. https://doi.org/10.1111/mec.12707

Mueller, J. C., Hermisson, J., Olano-Marín, J., Hansson, B., & Kempenaers, B. (2011). Linking genetic mechanisms of heterozygosity-fitness correlations to footprints of selection at single loci. Evolutionary Ecology, 25(1), 1–11. https://doi.org/10.1007/s10682-010-9377-2

Mueller, J. C., Kuhl, H., Timmermann, B., & Kempenaers, B. (2016). Characterization of the genome and transcriptome of the blue tit Cyanistes caeruleus: Polymorphisms, sex-biased expression and selection signals. Molecular Ecology Resources, 16(2), 549–561. https://doi.org/10.1111/1755-0998.12450

Nilsson, J. A., & Smith, H. G. (1988). Effects of dispersal date on winter flock establishment and social dominance in marsh tits Parus palustris. Journal of Animal Ecology, 57(3), 917–928. https://doi.org/10.2307/5101

Olano-Marín, J., Dawson, D. H., Girg, A., Hansson, B., Ljungvist, M., Kempenaers, B., & Mueller, J. C. (2010). A genome-wide set of 106 microsatellite markers for the blue tit (Cyanistes caeruleus). Molecular Ecology Resources, 10(3), 516–532. https://doi.org/10.1111/j.1755-0998.2009.02777.x

Olano-Marín, J., Mueller, J. C., & Kempenaers, B. (2011a). Heterozygosity and survival in blue tits (Cyanistes caeruleus): Contrasting effects of presumably functional and neutral loci. Molecular Ecology, 2019), 4028–4041. https://doi.org/10.1111/j.1365-294X.2011.05177.x

Olano-Marín, J., Mueller, J. C., & Kempenaers, B. (2011b). Correlations between heterozygosity and reproductive success in the blue tit (Cyanistes caeruleus): An analysis of inbreeding and single-locus effects. Evolution, 65(11), 3175–3194. https://doi.org/10.1111/j.1558-5646.2011.01369.x

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. PLoS Genetics, 2(12), 2071-2093. https://doi.org/10.1371/journal.pgen.0020190

Postma, E., & van Noordwijk, A. J. (2005). Gene flow maintains a large genetic difference in clutch size at a small spatial scale. Nature, 433, 65–68. https://doi.org/10.1038/nature03083

Pullido, F. (2007). The genetics and evolution of avian migration. BioScience, 57(2), 165-174. https://doi.org/10.1641/B570211

R Core Team. (2018). R: A language and environment for statistical computing, version 3.5.2. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Schlicht, L., & Kempenaers, B. (2015). Immediate effects of capture on nest visits of breeding blue tits, Cyanistes caeruleus, are substantial.