Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone

Ulrich Knief1,2, Christen M. Bossu2,3,4,9, Nicola Saino5, Bengt Hansson6, Jelmer Poelstra2,7, Nagarjun Vijay2,8, Matthias Weissensteiner1,2 and Jochen B. W. Wolf1,2*

The evolution of genetic barriers opposing interspecific gene flow is key to the origin of new species. Drawing from information on over 400 admixed genomes sourced from replicate transects across the European hybrid zone between all-black carrion crows and grey-coated hooded crows, we decipher the interplay between phenotypic divergence and selection at the molecular level. Over 68% of plumage variation was explained by epistasis between the gene NDP and a -2.8-megabase region on chromosome 18 with suppressed recombination. Both pigmentation loci showed evidence for divergent selection resisting introgression. This study reveals how few, large-effect loci can govern prezygotic isolation and shield phenotypic divergence from gene flow.

Understanding the origin of species has been foundational to the field of evolutionary biology. With the application of genomic approaches in natural populations, uncovering the genetic basis of speciation has come within reach8,9. All-black carrion crows (Corvus (corone) corone) and grey-coated hooded crows (C. (c.) cornix) meet in a stable and narrow contact zone that probably formed in the Early Holocene10. Assortative mating5 and social marginalization of minority phenotypes based on plumage pigmentation patterns act as evolutionary forces in maintaining phenotypic identity8,9. A screen of genomic outlier loci from information of admixed genomes to fine-map the genetic basis of speciation has come within reach1,2. All-black carrion crows and grey-coated hooded crows, we sampled a total of 409 individuals from replicate transects across the European hybrid zone (n = 236) (Supplementary Figs. 1 and 2) and further included individuals from allopatric populations to infer ancestry (central: n = 75; southern: n = 38). All individuals were genotyped for a final set of 1,111 single nucleotide polymorphisms (SNPs) selected from 16.6 million variants segregating in European populations, including highly ancestry-informative (high Pst: n = 735; fixed: n = 51) and background markers (n = 325; Supplementary Fig. 3 and Supplementary Table 1). Markers were equally spread across the genome except for chromosome 18, which was densely covered with 230 tightly linked markers in the previously identified outlier region5 (median linkage disequilibrium: central r = 0.18; southern r = 0.17). For a set of downstream analyses, chromosome 18 was treated separately from the remaining genome-wide SNPs (median linkage disequilibrium within chromosomes: r = 0.0024).

To map the genetic basis of phenotypic divergence, we first quantified plumage variation in 129 southern hybrids. We divided the dorsal and ventral plumage surface into discrete patches and scored shading on a grey scale reflecting the amount of eumelanin deposited into the feathering11 (Supplementary Fig. 4). Interindividual variation in shading was highly correlated among patches (average Spearman’s r = 0.76; Supplementary Fig. 5a). Next, we performed principal component analysis (PCA), in which 85.88% of the total phenotypic variance was captured by the first two components (PC1: 78.22%; PC2: 7.66%). Positive PC1 scores corresponded to a darker appearance (Fig. 1a), whereas PC2 separated the central body parts (belly and mantle) from distal and caudal parts (Supplementary Fig. 5b).

Genome-wide association (GWA) mapping uncovered one major genomic region strongly associated with variation in PC1. Out of a total of 170 SNPs remaining significant after multiple testing correction, 166 fell within ~2.8 Mb overlapping the genomic outlier region on chromosome 18 (chr18), with single SNPs explaining up to 76.07% (SNP: scaffold_78:1682876) of the variance in PC1 (median: 54.0%; Fig. 1b, Supplementary Fig. 6 and Supplementary Table 2). High linkage disequilibrium among SNPs in this region and low levels of interspecific recombination allowed categorization of this region by ancestry diplotypes (Fig. 1c): a purebred carrion crow type homoygous for the ‘dark’ (D) allele (chr18D), a purebred hooded crow type homoygous for the ‘light’ (L) allele (chr18L) and the heterospecific type (chr18H). These ancestry types predicted segregating phenotypic variation in PC1 equally well as the most

1Division of Evolutionary Biology, Faculty of Biology, LMU Munich, Munich, Germany. 2Science for Life Laboratories and Department of Evolutionary Biology, Uppsala University, Uppsala, Sweden. 3Population Genetics, Department of Zoology, Stockholm University, Stockholm, Sweden. 4Institute of the Environment and Sustainability, Center for Tropical Research, University of California, Los Angeles, CA, USA. 5Department of Environmental Science and Policy, University of Milan, Milan, Italy. 6Department of Biology, Lund University, Lund, Sweden. 7Biology Department, Duke University, Durham, NC, USA. 8Department of Biological Sciences, Indian Institute of Science Education and Research, Bhopal, India. 9These authors contributed equally: Ulrich Knief, Christen M. Bossu. *e-mail: j.wolf@biologie.uni-muenchen.de
significant SNPs (74.29 versus 76.07%, respectively). Chr18DD individuals (n = 29) were completely black, corresponding to a narrow distribution of high PC1 scores (s.d. in PC1 scores = 0.07; range: 3.96–4.33) (Fig. 1d). In contrast, chr18LL individuals had a much wider phenotypic distribution (s.d. = 1.46; range: −3.74–2.18): while the majority of individuals (n = 35) displayed hooded crow phenotypes, 14 individuals showed intermediate plumage patterns with extensive variation among them. However, the upper belly and mantle were grey in all cases. Individuals of the heterospecific chr18DL type covered almost the entire phenotypic spectrum except for the extremes of pure black or the grey-coated hooded phenotypes (s.d. in PC1 scores = 1.96; range: −3.34–3.59; Fig. 1d).

These results demonstrate that phenotypic variation was largely, but not exclusively, governed by a genetic factor within chr18 containing 88 genes. Additional environmental or genetic factors are required to explain the residual phenotypic variation within chr18DL and chr18LL types. Three SNPs in proximity to the genes NDP and EFHC2 on chromosome 1, and a single SNP in the first intron of LRP6 on chromosome 1A, explained an additional 10.27 and 0.43% of the variance in PC1, respectively (Supplementary Table 2). Norrin (NDP) and low-density lipoprotein receptors (LRP5/6) closely interact in the Wnt signalling pathway14 and have been hypothesized to be involved in phenotypic divergence across multiple independent crow hybrid zones15. Gene expression of NDP has further been associated with pigmentation patterning and divergence between carrion and hooded crows9. Recently, NDP has also been suggested to regulate melanin-based patterning in pigeons15. NDP, rather than EFHC2, is thus a prime candidate for modulating both the intensity and

---

Fig. 1 | Genetic basis of phenotypic variation. a, Examples of pure parental (CC, carrion crow; HC, hooded crow) and hybrid (Fx) phenotypes. Plumage pigmentation was scored on a grey scale for dorsal (D) and ventral (V) patches, and subsequently summarized by PCA. The abscissa shows the smoothed density distribution of PC1 scores for 122 individuals, with the underlying shading for orientation. b, Manhattan plot of GWA analysis for PC1 scores shown for 1,103 SNPs by chromosome, including SNPs located on unassigned scaffolds (Un) and the sex chromosome (Z). Significantly associated SNPs are shown in red. The dashed line represents the genome-wide significance threshold. c, Left, close-up of the 74 most ancestry-informative SNPs within chr18. SNPs are coloured by ancestry of the parental populations (black, carrion crow; light grey, hooded crow; dark grey, heterospecific). Right, ancestry components inferred across all SNP group individuals into three discrete classes corresponding to the three diplotypes: chr18DD, chr18DL and chr18LL. d, Decomposition of phenotypic variation as predicted by the recessive, epistatic interaction between the gene NDP and chr18 ancestry type. Alleles are named by their inferred phenotypic effect (D, dark; L, light).
position of plumage pigmentation, consistent with the wide phenotypic distribution in chr18\textsubscript{LL} and chr18\textsubscript{DL} individuals (Fig. 1d). Next, we included dominance and epistasis in the statistical model. The best model accounted for 87.91\% of the variance in PC1, corresponding to $78.22\times87.91=68.76\%$ of the total phenotypic variance. It supported additive and dominance effects in PC1, corresponding to 78.22\% of the variance explained by the model. The best model accounted for 87.91\% of the variance explained by the model.

chr18\textsubscript{LL} position of plumage pigmentation, consistent with the wide phenotypic distribution across the hybrid zone. To quantify selection acting on phenotypic variance. It supported additive and dominance effects in PC1, corresponding to 78.22\% of the variance explained by the model. The best model accounted for 87.91\% of the variance explained by the model.

chr18\textsubscript{DL} and chr18\textsubscript{LL} (Fig. 1d). Individuals of both diploid genotypes (chr18\textsubscript{DL} and chr18\textsubscript{LL}) became increasingly lighter in combination with the NDP\textsubscript{DL} and NDP\textsubscript{LL} genotype. Only chr18\textsubscript{LL} individuals homozygous for the light NDP allele (that is, NDP\textsubscript{LL}) recovered the hooded crow phenotype.

All three major loci—chr18, NDP and LRP6—resided within the few previously identified genomic regions of increased differentiation between parental populations\textsuperscript{14,15}. Moreover, the 170 significant SNPs identified in the GWA study were among those having the highest $F_{ST}$ values (generalized linear mixed model: $P$ value $<2\times10^{-15}$; Supplementary Fig. 7). $F_{ST}$ outliers need not be associated with the genes underlying phenotypic divergence, in particular for polygenic trait architectures\textsuperscript{16}. The association of all genetic factors controlling phenotypic variation with genetic differentiation among pure, parental populations thus constitutes a first indication that the loci controlling phenotypic divergence may themselves be subject to divergent selection limiting gene flow across the hybrid zone. To quantify selection acting on admixed genomes segregating in the contemporary hybrid zone, we performed clinal analyses. First, we considered gene flow along a spatial axis across the hybrid zone. Transition of genome-wide ancestry across both geographic transects was best explained by a sigmoid clinal function (Fig. 2a,c and Supplementary Table 3), supporting previous evidence for recurrent backcrossing\textsuperscript{1}. Inferred cline widths of 316.43 km (central) and 553.85 km (southern) substantially exceeded morphologically based estimates of 45–100 km (central) and 10–100 km (southern) (Supplementary Table 4). This provides further evidence for genome-wide introgression expanding far beyond morphologically inferred boundaries\textsuperscript{16}. In contrast to neutral genetic variation, barrier loci under divergent selection favor introgression and are thus expected to show steeper clines and reduced width while maintaining the centre of the clines\textsuperscript{12}. Consistent with this prediction, chr18 and NDP shared cline centres with genome-wide estimates, but were significantly reduced in width (Fig. 2b,f, Supplementary Fig. 8 and Supplementary Table 5). Estimates of 58 km (chr18) and 105 km (NDP) in the central zone closely mimicked morphology-based inference (Supplementary Table 4). Similarly, narrow clines were observed in the southern contact zone for NDP (65 km) and chr18 (Fig. 2f, Supplementary Fig. 8 and Supplementary Table 5), where the formal estimate of 241.35 km was probably inflated due to incomplete sampling of populations with carrion crow ancestry (Fig. 2f).

Genomic clinal analysis provides an alternative way to infer recent selection against hybridization. It does not require a spatial axis, but describes locus-specific introgression along a gradient of genome-wide admixture\textsuperscript{17}. Loci subject to divergent selection are associated with lower fitness in heterospecific genomic backgrounds and are

Fig. 2 | Geographic and genomiccline analyses. a,b,f. Geographic clines for hybrid indices estimated from genome-wide data (a and e; $n=752$ SNPs) and SNPs on chromosome 18 (b and f; $n=230$ SNPs). Estimates are shown for both the central hybrid zone (a and b; $n=116$ individuals) and southern hybrid zone (e and f; $n=273$ individuals). Depicted are the maximum-likelihood clines and observed average hybrid indices per sample location (with the 95\% credible cline region shaded in grey). Each cline extends from the allopatric carrion to the allopatric hooded crow populations. Circle areas reflect sample sizes, colours are used for orientation (dark grey, carrion crow; light grey, hooded crow). The abscissa is centred on the cline centre. Rectangles (yellow, central; red, southern) depict the cline widths predicted by the models. c,d,g,h. Genomic clines for all $1,111$ SNPs across the genome (c and g) and for the outlier region on chromosome 18 (chr18) specifically (d and h). Summary statistics are shown for the central (c and d) and southern hybrid zone (g and h). Depicted is the $-\log_{10}$[P value] of the cline rate $v$, with high values reflecting departures of introgression from the genome average. SNPs with the strongest evidence (upper 5 percentiles) for reduced introgression are coloured (yellow, central; red, south). The dashed lines in g and h represent the genome-wide significance threshold.
that genes controlling phenotypic variation are subject to divergent variation is impeded by high linkage disequilibrium owing to low recombination, possibly due to an inversion or proximity to the centromere. Nevertheless, AXIN2, located under the first FST peak, is a prime candidate. Like NDP, it acts in the Wnt pathway and is involved in pigmentation and pattern formation. PRKCA and a tandem array of CACNG genes act in the mitogen-activated protein kinase pathway and reside in the second FST peak. This region was also characterized by signatures of positive selection in the lineage leading to the hooded crow. Variation in hybrid colour phenotypes was best attributed to recessive epistatic interaction between AXIN2 and PRKCA/CACNG as responsible for phenotypic divergence.

This study demonstrates the power of leveraging information from naturally occurring hybridization for speciation research. In combination with outlier scans of parental populations, admixture analyses provide valuable insights into the genetic architecture of hybridization, its genetic control, and selection acting on it. Variation in hybrid colour phenotypes was best attributed to recessive epistatic interaction between AXIN2 and PRKCA/CACNG as responsible for phenotypic divergence.

Sample preparation and SNP genotyping. We extracted DNA from blood samples using a standard phenol chloroform assay, and from muscle tissue using the DNeasy blood and tissue kit (Qiagen). DNA quantity and quality were assessed using Nanodrop and the SYBR Green fluorescence assay (Invitrogen). Polymorphic diagnostic SNPs were selected in the hooded crow populations of the central hybrid zone near Graz. Individuals allopatric to the central hybrid zone were collected in north-western Germany (adult hooded crows; n = 45) and Poland and Sweden (adult hooded crows; n = 30). For all of these individuals, blood samples were taken from the brachial vein and stored either in Queen’s lysis buffer, or ethylenediaminetetraacetic acid- or heparin-coated sample containers.

In the southern hybrid zone, adult birds were shot by local hunters and tissue samples from the breast muscle were stored in ethanol (n = 238 individuals, including 107 samples from ref. 3 and 131 samples taken by N.S. for this study). Additionally, we obtained tissue samples from naturally occurring hybridization for speciation research11. In the central hybrid zone, we collected 152 nestlings from 76 nests in May–June 2007, 2008, 2013 and 2014. Additionally, we obtained samples from 26 nestlings that were raised in captivity and originated from the central hybrid zone near Graz. Individuals allopatric to the central hybrid zone were collected in north-western Germany (adult hooded crows; n = 45) and Poland and Sweden (adult hooded crows; n = 30). For all of these individuals, blood samples were taken from the brachial vein and stored either in Queen’s lysis buffer, or ethylenediaminetetraacetic acid- or heparin-coated sample containers.

In the southern hybrid zone, adult birds were shot by local hunters and tissue samples from the breast muscle were stored in ethanol (n = 238 individuals, including 107 samples from ref. 3 and 131 samples taken by N.S. for this study). Additionally, we obtained tissue samples from naturally occurring hybridization for speciation research11. In the central hybrid zone, we collected 152 nestlings from 76 nests in May–June 2007, 2008, 2013 and 2014. Additionally, we obtained samples from 26 nestlings that were raised in captivity and originated from the central hybrid zone near Graz. Individuals allopatric to the central hybrid zone were collected in north-western Germany (adult hooded crows; n = 45) and Poland and Sweden (adult hooded crows; n = 30). For all of these individuals, blood samples were taken from the brachial vein and stored either in Queen’s lysis buffer, or ethylenediaminetetraacetic acid- or heparin-coated sample containers.

In the southern hybrid zone, adult birds were shot by local hunters and tissue samples from the breast muscle were stored in ethanol (n = 238 individuals, including 107 samples from ref. 3 and 131 samples taken by N.S. for this study). Additionally, we obtained tissue samples from naturally occurring hybridization for speciation research11. In the central hybrid zone, we collected 152 nestlings from 76 nests in May–June 2007, 2008, 2013 and 2014. Additionally, we obtained samples from 26 nestlings that were raised in captivity and originated from the central hybrid zone near Graz. Individuals allopatric to the central hybrid zone were collected in north-western Germany (adult hooded crows; n = 45) and Poland and Sweden (adult hooded crows; n = 30). For all of these individuals, blood samples were taken from the brachial vein and stored either in Queen’s lysis buffer, or ethylenediaminetetraacetic acid- or heparin-coated sample containers.
candidate genes, we allowed one variant with the highest primer design score in each gene region (for example, the exon, intron or regulatory region, defined as 3 kb surrounding a gene), which resulted in a higher density within the peak region on chromosome 18.

The exact location of each SNP with reference to genome version 2.5 (ReSeq Assembly ID according to the National Center for Biotechnology Information: GCF_000738375.1, as published in ref. 1) and primer sequences are available in Supplementary Table 1. GoldenGate genotyping was performed on 520 samples at the SNP&SEQ Technology Platform, Uppsala. Cluster plots were automatically analysed using Illumina's GenomeStudio software (version 2011.1). We removed 31 individuals that had a call rate of 0%, and 2 additional duplicated individuals. We further removed 41 SNPs that had a missing call rate > 25% (n = 37) or were monomorphic (n = 4). This resulted in a final set of 51 fixed, 735 outlier and 325 background genome-wide SNPs. For the fixed set, FST ranged from 0.766–1.000 (mean = 0.897) and the background set from 0.83–1.000 (mean = 0.970) between allopatric populations in the central and southern hybrid zone, respectively. For the outlier set, it ranged from −0.022–1.000 (mean = 0.199) and −0.050–1.000 (mean = 0.282), and for the background set from −0.022–0.378 (mean = 0.029) and −0.047–0.574 (mean = 0.100), respectively. One individual was genotyped three times to assess genotyping errors. No discordant genotype calls were observed among 2,217 comparisons, suggesting an accuracy of genotype calls above 99.95%. Of the allopatric populations, 40 individuals had previously been genotyped using the HaploTypeCaller in GATK (version 3.3.0)2 after paired-end whole-genome sequencing on the HiSeq2000 (Illumina) platform. Coverage ranged from 6.83x to 23.45x, median = 16.66x. For those individuals, we used the parallelnewhybrid (version 2.07), we further estimated the posterior probability of an individual belonging to the background set from the southern hybrid zone, respectively, removing nest mates; Supplementary Table 9). We call these clear diploids based on chromosome 18 chr18LL, chr18LR and chr18UR, respectively.

We additionally used these plots to identify interspecific recombination breakpoints along chromosome 18 by taking change in haplotype ancestry as evidence for a recombination event (Supplementary Fig. 13). We only considered changes in haplotype ancestry at lethal or neutral SNPs, allowing a mutation rate of at least three mutations per million years. Breakpoints along chromosome 18 were either classified as purebreds (65.8 and 72.9% in the central and southern hybrid zone, respectively) or as F1 hybrids (30.3 and 25.8% in the central and southern hybrid zone, respectively). Using this cut-off, we estimated introgress’ hybrid indices were in good correspondence using the SNPs on chromosome 18 (Supplementary Fig. 12). We visually inspected graphical representations of marker ancestry across chromosome 18 in all F1-hybrid individuals (using introgress’ mk.image() function), which showed that these individuals were indeed heterogeneous at the outlier loci on chromosome 18. Using the SNPs on chromosome 18, NewHybrids assigned all individuals to the genealogical classes with a posterior probability of 1, and most of the individuals were either classified as purebreds (65.8 and 72.9% in the central and southern hybrid zone, respectively) or as F1 hybrids (30.3 and 25.8% in the central and southern hybrid zone, respectively). Assuming randomness of crossovers, haplotype tracts of the exact same lengths shared between nest mates are likely to be identical by descent and were only counted as a single recombination event.

Phenotype characterization and genome-wide associations. Digital photographs of the ventral and dorsal side were taken for all birds from Graz and around half of the birds from the southern hybrid zone (n = 109 pictures of 18 individuals from Graz and n = 456 pictures of 111 individuals from the southern hybrid zone) for scoring the plumage. For those individuals, we scored the amount of grey and black in the feathering (Fig. 1a and Supplementary Fig. 4). Each patch was scored as 0 = pure grey, 1 = dark grey or a mixture of grey and black feathers or 2 = pure black by the same person (U.K.; Supplementary Table 10). Because some plumage patches could not be scored reliably, we imputed missing data (2.18%) using the mice algorithm (version 1.11)19. Individual measurements were then summarized using PCA, as implemented in the FactoMineR R package (version 1.39)20. All 11 variables loaded positively on PC1, which explained 78.22% of the variance in pigmentation pattern and intensity. PC2 explained 7.66% of the variance and separated the belly and mantle from the other plumage patches. To assess the objectivity of scoring, we estimated interobserver repeatability of the colour PC1 and PC2 scores by getting measurements taken by a second person (N.S.) on a subset of individuals (n = 105 individuals). We estimated the repeatability of PC1 and PC2 scores using the rptGaussian() function of the rprR package (version 0.9.21)21. Both PC1 and PC2 scores were highly repeatable between observers (PC1: r.s.e. = 0.965 ± 0.007; P = 4 × 10−4*; PC2: r.s.e. = 0.646 ± 0.057; P = 1 × 10−4*) using n = 10.000 parametric bootstrap intervals for estimation). We performed two GWA studies, using either colour PC1 or colour PC2 as dependent variables and each SNP as a covariate (coded as 0, 1 or 2 copies of the minor allele) using 1 degree of freedom (additive effect). Models were fitted with the gwas() function and a Gaussian error structure in the GenABEL R package (version 1.8−0)22. Since SNPs showed varying degrees of linkage disequilibrium, we used the simpleM-algorithm with default settings to estimate the effective number of tests performed23 and used this estimate to control the genome-wide type I error rate.

We fitted all 18 SNPs showing a significant additive main effect on colour PC1 (3 SNPs on chromosome 1, 1 SNP on chromosome 1A and 166 SNPs on chromosome 18) as factors to test for additional dominant gene action, and tested epistatic interactions between all of them (3 × 1 + 3 × 166 + 1 × 166 = 667 interactions). For significance testing, we performed likelihood ratio tests comparing models with and without the dominant gene action or interaction. Because both chromosomes 18 and the three SNPs on chromosome 1 effectively behaved as one locus, we reduced our model to include only the ancestry on chromosome 18 and the most significant SNP on chromosome 1 in our final model. We selected the best model based on Akaike’s information criterion using ΔAIC ≥ 2 as the selection threshold.24

For each of the four simulated datasets (4× central autosomal SNPs, 4× southern autosomal SNPs, 4× central chromosome 18 SNPs and 4× southern chromosome 18 SNPs) and estimated NewHybrids’ assignment efficiency, accuracy and overall performance. We used the simulated data to select the optimal posterior probability cut-off that maximized the accuracy of NewHybrids genealogical class assignment. The maximum was at an overall posterior probability of 0.95 covering all genealogical classes for both SNP sets in the two hybrid zones. Using this cut-off, NewHybrids assigned 99.14% of purebreds, 85.98% of F1, 85.75% of F2 and 95.98% of backcrosses correctly.

For the empirical data, NewHybrids’ genealogical class assignments and introgress’ hybrid indices were in good correspondence using the SNPs on chromosome 18 (Supplementary Fig. 12). We visually inspected graphical representations of marker ancestry across chromosome 18 in all F1-hybrid individuals (using introgress’ mk.image() function), which showed that these individuals were indeed heterogeneous at the outlier loci on chromosome 18. Using the SNPs on chromosome 18, NewHybrids assigned all individuals to the genealogical classes with a posterior probability of 1, and most of the individuals were either classified as purebreds (65.8 and 72.9% in the central and southern hybrid zone, respectively) or as F1 hybrids (30.3 and 25.8% in the central and southern hybrid zone, respectively). Assuming randomness of crossovers, haplotype tracts of the exact same lengths shared between nest mates are likely to be identical by descent and were only counted as a single recombination event.

Geographic clines. We used one-dimensional geographic cline analysis to assess whether genetic variation in the outlier region of chromosome 18 and the significant loci from the GWA study on chromosomes 1 and 1A showed a signal of contemporary divergent selection. Several sampling locations did not fully coincide with the transect line. In these cases, we collapsed the two-dimensional geographic coordinates using PCA for both the central and southern hybrid zone
These likely represent ancestral polymorphisms shared between American crows, hooded and carrion crows and are thus uninformative with respect to the ancestral state in hooded and carrion crows. Some 3.23% of all sites were fixed for the same allele in American crows and rooks, but polymorphic in jackdaws. This could be due to either a mutation in the lineage leading to jackdaws (homoplasly) or incorrectly mapped paralogues, and these sites were thus excluded. We also removed 1.88% of the sites that were fixed for alternative alleles in two species and polymorphic in the third or biallelic in two or more species. Finally, we discarded sites that were polymorphic in rooks and fixed for the same allele in jackdaws and American crows (1.43%), those that were triallelic (0.45%) or those where genotype information was missing (0.45%). For a summary of the data, see Supplementary Table 11.

F statistics estimation. \( F_{st} \) values were estimated from genotypes derived from whole-genome sequencing data of the allopatric populations’ using PLINK (version 1.90b4.4). We used the allopatric samples of carrion and hooded crows from the central and southern hybrid zone independently as input populations. Wright’s inbreeding coefficient (\( F_{st} \)) values were estimated from chromosome 18 genotypes of individuals sampled in the central and southern hybrid zone that were assigned either a hooded or carrion crow or F1-hybrid ancestry. For this, we used the R package hierfstat (version 0.04-22).

PCA. To better understand the partitioning of genetic variation along chromosome 18 (see refs. 30, 31), we performed PCA of genotyping variation using all individuals sampled in the two hybrid zones (\( n = 236 \) individuals for the southern zone) and all SNPs on chromosome 18 for which we could resolve the ancestral state and that had a missing call rate smaller than 0.05 and a minor allele frequency larger than 0.05 (\( n = 204 \) SNPs for the central zone and \( n = 193 \) SNPs for the southern zone). Discrete clusters of genetic variation provide evidence for divergent haplotypes that are non-recombining (for example, due to an inversion). In the case of two major haplotypes, we expected PCA to classify individuals by diploidy: homozgyous individuals are expected to cluster at both ends of the PCA distribution with heterozygous individuals in between. We further expect the squared correlation coefficient of PCA loadings with the chr18 diplotype to reflect \( F_{st} \) (termed communality \( h^2 \)) if the haplotype structure on chromosome 18 was the same between the allopatric populations (used for \( F_{st} \)) and the individuals from the hybrid zone. The ancestry information for all SNPs enabled us to polarize the principal component loadings. This provides a directional measure of population differentiation, which should provide information on the direction of positive selection (increase in the proportion of derived variants in the target population). PCA was performed using the R package SNPRelate (version 1.12.1).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

Genotype and phenotype data are available in Supplementary Tables 1, 8, 10 and 11. R scripts used for the analyses are available as Supplementary Data 1–7.

**Received:** 9 October 2018; **Accepted:** 18 February 2019; **Published online:** 25 March 2019

**References**

1. Wolf, J. B. W. & Ellegren, H. Nat. Rev. Genet. 18, 87–100 (2017).
2. Ravinet, M. et al. J. Evol. Biol. 30, 1450–1477 (2017).
3. Vijay, N. et al. Nat. Commun. 7, 13195 (2016).
4. Mayr, E. Systematics and the Origin of Species (Columbia Univ. Press. 1942).
5. Randler, C. Ardea 95, 143–149 (2007).
6. Saino, N. & Scatizzi, L. Boll. Zool. 58, 255–260 (1991).
7. Londei, T. Ibis 155, 632–634 (2013).
8. Poelstra, J. W. et al. Science 344, 1410–1414 (2014).
9. Poelstra, J. W., Vijay, N., Hoopner, M. P. & Wolf, J. B. W. Mol. Evol. 24, 4617–4628 (2015).
10. Le Corre, V. & Kremer, A. Mol. Ecol. 21, 1548–1566 (2012).
11. Gottard, P., Mandeville, E. G. & Buerkle, C. A. Annu. Rev. Ecol. Evol. Syst. 48, 207–229 (2017).
12. Barton, N. H. & Hewitt, G. M. Annu. Rev. Ecol. Syst. 16, 113–148 (1985).
13. Wu, Y. C. et al. In situ quantification of individual mRNA transcripts in melanomaculose discloses gene regulation of relevance to specialization. J. Exp. Biol. https://doi.org/10.1242/jeb.194431 (2019).
14. Ke, J. Y. et al. Gene Dev. 27, 2305–2319 (2013).
15. Vickrey, A. J. et al. eLife 7, e34803 (2018).
16. Haas, F. et al. Mol. Ecol. Resour. 18, 304–315 (2018).
17. Sivaraman, J. M. & Barton, N. H. Evolution 40, 1141–1159 (1986).
18. Weissensteiner, M. H. et al. Genome Res. 27, 697–708 (2017).
19. Kofron, M. et al. Dev. Biol. 237, 183–201 (2001).
20. Duquet, M. Ornithos 19, 57–67 (2012).
Acknowledgements

This study would not have been possible without the commitment of dedicated ornithologists who helped to locate active nest sites and participated in sampling. These include M. Hug and colleagues in Brandenburg, J. Voigt, D. Kronbach, J. Wollmerstadt, M. Schrack and W. Nachtigall in Sachsen, M. Döpfner in Baden-Württemberg, S. Zinko and M. Grossmann in Austria, and personnel of the Amministrazione Provinciale of Alessandria, Asti and Canoe in Italy. We further acknowledge the Max Planck Institute for Ornithology in Radolfzell, the Friedrich-Loßler-Institute and the Förderverein Sächsische Vogelschutzwarte Neschwitz e. V. for help with the organization of field work. The UPPMAX Next-Generation Sequencing Cluster and Storage (UPPNEX) project, funded by the Knut and Alice Wallenberg Foundation and Swedish National Infrastructure for Computing, provided access to computational resources. Funding was provided by the Volkswagentiftung (grant U/83 496 to J.B.W.W.), European Research Council (ERC StG-336536 FuncSpecGen to J.B.W.W.), Knut and Alice Wallenberg Foundation (project grant including J.B.W.W.) and LMU Munich (to J.B.W.W.).

Author contributions

U.K., C.M.B. and J.B.W.W. conceived of the study design. C.M.B., J.P., M.W., B.H., N.S. and J.B.W.W. conducted the field work and provided samples. C.M.B. and U.K. performed all analyses. N.S. and J.P. helped with phenotype scoring, and N.V. assisted in SNP design. U.K., C.M.B. and J.B.W.W. wrote the manuscript with input from all other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-019-0847-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.B.W.W.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019
# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| Item                                                                 | Confirmed |
|----------------------------------------------------------------------|-----------|
| n/a                                                                  |           |
| - The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement |           |
| - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |           |
| - The statistical test(s) used AND whether they are one- or two-sided |           |
| - Only common tests should be described solely by name; describe more complex techniques in the Methods section. |           |
| - A description of all covariates tested                            |           |
| - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |           |
| - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |           |
| - For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted Give P values as exact values whenever suitable. |           |
| - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |           |
| - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |           |
| - Estimates of effect sizes (e.g. Cohen's $d$, Pearson’s $r$), indicating how they were calculated |           |

Our web collection on statistics for biologists contains articles on many of the points above.

## Software and code

**Policy information about availability of computer code**

- **Data collection**: We performed GoldenGate (Illumina) genotyping. The resulting cluster plots were automatically analysed using Illumina's GenomeStudio software (v2011.1).
- **Data analysis**: We used publicly available software for all analyses: R (v3.4.3) and specific packages as named throughout the manuscript (with their version numbers). We further use the PLINK (v1.90b4.1), NewHybrids (v2.0+ Developmental, July/August 2007) and bgc (v1.03) software.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

**Policy information about availability of data**

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Genotype and phenotype data are available in Supplementary Tables 1, 8, 10 and 11. R-scripts used for the analyses are available as Supplementary Data 1 to 7.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences
Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description
Drawing from information on over 400 admixed genomes in replicate transects across the hybrid zone between all-black carrion crows and grey-coated hooded crows, we shed light on the interplay between phenotypic divergence and selection at the molecular level. For that we use whole-genome genotyping and perform cline analyses and a GWAS.

Research sample
522 carrion and hooded crows and their hybrids that were sampled across the hybrid zone (in Germany and Italy) and from allopatric populations in Germany, Italy, Sweden and Poland.

Sampling strategy
In Germany, we sampled blood from nestlings. In Italy, adult birds were shot by local hunters.

Data collection
CB, JP, MW, BH, NS and JW conducted field work and provided samples.

Timing and spatial scale
May–June 2007, 2008, 2013 and 2014.

Data exclusions
No data was excluded and all analyses were performed blind in respect to the outcome.

Reproducibility
Data were collected across the hybrid zone in two independent locations and similar results were obtained in both.

Randomization
Not applicable in our study.

Blin ding
All analyses were performed blind in respect to the outcome.

Field work, collection and transport

Field conditions
Not relevant for the current study because we focus on genetic data.

Location
We obtained blood and tissue samples of carrion and hooded crows (Corvus (corone) corone and C. (c.) cornix) and their hybrids along two transects across the European hybrid zone (Supplementary Figs. 1 and 2). Transects were chosen such that they included phenotypically pure populations resembling the parental allopatric populations at the endpoints, and several geographically spaced populations with mixed hybrid phenotypes in between. One transect was located in eastern Germany (and Graz, Austria) and the other at the north-western border of Italy towards France. We hereafter refer to these transects as “central” and “south”, respectively. In the central hybrid zone, we collected 152 nestlings from 76 nests in May–June 2007, 2008, 2013 and 2014. Additionally, we obtained samples from 26 nestlings that were raised in captivity and originated from the central hybrid zone near Graz. Individuals allopatric to the central hybrid zone were caught in north-western Germany (adult carrion crows, N = 45) and Poland and Sweden (adult hooded crows, N = 30).

In the southern hybrid zone, adult birds were shot by local hunters and tissue samples from the breast muscle were stored in Ethanol (N = 238 individuals including 107 samples from ref16 and 131 samples taken by N.S. for this study). Allopatric populations of the southern transect consisted of 26 adult carrion crows from southern Germany and 12 adult hooded crows from central Italy, for all of which blood samples were taken. For more information on samples and sampling locations consult Supplementary Tables 7 and 8.

Access and import/export
Permissions for sampling of wild crows were granted by Regierungspräsidium Freiburg (Aktenzeichen: 55-8852.15), Landratsamt Zwickau (364.622-N-Her-1/14), Landratsamt Mittelsachsen (55410704 Beriungserf-Voigt_14), Landratsamt Vogtlandkreis (364.622-2-2-88841/2014), Landratsamt Meißen (672/364.621-Kennzeichnung von Tieren-18935/2013), Landratsamt Bauzen (67.3-364.622:13-01-Krähen), Landesdirektion Sachsen (24-9168.00/2013-4), Landesamt für Verbraucherschutz, Landwirtschaft und Flurneuordnung Brandenburg (23-2347-8a182008) in Germany and by Jordbruksverket (Dnr 30-1326/10) in Sweden. Polish hooded crow nestlings were provided by courtesy of Dr. Andrzej Kruzewicz from the animal rehabilitation centre of Warsaw Zoo. Italian hooded crows of the allopatric population were provided by Centro Recupero Fauna Selvatica LIPU di Roma, Rome, Italy. Crow specimens from the southern hybrid zone were provided by the local administrations (Province di Cuneo ed Alessandria) in the frame of annual spring shooting of bird pests.

Disturbance
Samples in the German hybrid zone were taken from nestlings and they were released after handling. Crow specimens from the southern hybrid zone were provided by the local administrations in the frame of annual spring shooting of bird pests.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.
Animals and other organisms

Policy information about studies involving animals, ARRIVE guidelines recommended for reporting animal research

| Laboratory animals | NA |
|--------------------|----|
| Wild animals       | Samples in the German hybrid zone were taken from nestlings (N=178). Crow specimens from the southern hybrid zone were provided by the local administrations in the frame of annual spring shooting of bird pests (N=238). |
| Field-collected samples | NA |
| Ethics oversight   | Permissions for sampling of wild crows were granted by Regierungspräsidium Freiburg (Aktenzeichen: 55-8852.15), Landratsamt Zwickau (364.622-N-Her-7/14), Landratsamt Mittelsachsen (55410704 Berichtigungserl-Voigt_14), Landratsamt Vogtlandkreis (364.622-2-2-88841/2014), Landratsamt Meißen (672/364.621-Kennzeichnung von Tieren-18935/2013), Landratsamt Bauen (67.3-364.622:13-01-Krähen), Landesdirektion Sachsen (24-9168.00/2013-4), Landesamt für Verbraucherschutz, Landwirtschaft und Flurneuordnung Brandenburg (23-2347-8a182008) in Germany and by Jordbruksverket (Dnr 30-1326/10) in Sweden. Polish hooded crow nestlings were provided by courtesy of Dr. Andrzej Kruszewicz from the animal rehabilitation centre of Warsaw Zoo. Italian hooded crows of the allopatric population were provided by Centro Recupero Fauna Selvatica LIPU di Roma, Rome, Italy. Crow specimens from the southern hybrid zone were provided by the local administrations (Province di Cuneo ed Alessandria) in the frame of annual spring shooting of bird pests. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.