

Published in final edited form as:

Nat Ecol Evol. 2019 April ; 3(4): 570–576. doi:10.1038/s41559-019-0847-9.

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

Ulrich Knief^{#1}, Christen M. Bossu^{#2,3,4}, Nicola Saino⁵, Bengt Hansson⁶, Jelmer Poelstra^{2,7}, Nagarjun Vijay^{2,8}, Matthias Weissensteiner^{1,2}, and Jochen B. W. Wolf^{1,2,*}

¹Division of Evolutionary Biology, Faculty of Biology, LMU Munich, 82152 Planegg-Martinsried, Germany

²Science for Life Laboratories and Department of Evolutionary Biology, Uppsala University, 75236 Uppsala, Sweden

³Department of Zoology, Population Genetics, Stockholm University, SE-106 91 Stockholm, Sweden

⁴Institute for the Environment and Sustainability, Center for Tropical Research, University of California, CA 90095 Los Angeles, US

⁵Department of Environmental Science and Policy, University of Milan, I-20133 Milan, Italy

⁶Department of Biology, Lund University, SE-223 62 Lund, Sweden

⁷Biology Department, Duke University, NC 27708 Durham, US

⁸Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Madhya Pradesh, India

These authors contributed equally to this work.

Abstract

The evolution of genetic barriers opposing inter-specific gene flow is key to the origin of new species. Drawing from information of over 400 admixed genomes sourced from replicate transects

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence and requests for materials should be addressed to JW. Jochen Wolf, Division of Evolutionary Biology, Faculty of Biology, Ludwig Maximilians University Munich, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany, Phone: 0049-89-2180-74102, Fax: 0049-89-2180-74104, j.wolf@biologie.uni-muenchen.de.

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

Data and computer code 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.

Author contributions

UK, CB and JW conceived of the study design. CB, JP, MW, BH, NS and JW conducted field work and provided samples. CB and UK performed all analyses with the help from NS and JP for phenotype scoring and aid in SNP design by NV. UK, CB and JW wrote the manuscript with input from all other authors.

Competing interests

The authors declare no competing interests.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

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 Mb 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 reach^{1,2}. All-black carrion crows (*Corvus (corone) corone*) and grey-coated hooded crows (*C. (c.) cornix*) meet in a stable and narrow contact zone that has likely formed in the early Holocene^{3,4}. Assortative mating⁵ and social marginalization of minority phenotypes⁶ based on plumage pigmentation patterns act in concert to reduce the amount of crossbreeding⁷. Genome scans in the European hybrid zone suggest that genetic variation is homogenized genome-wide by inter-specific gene flow beyond the morphologically visible confines of the hybrid zone⁸. A few genomic regions, however, including a ~2.8 Mb region on chromosome 18 with low levels of recombination, have been identified as candidate barrier loci that resist gene flow and might contribute to maintaining phenotypic identity^{8,9}. A screen of genomic outlier loci at independent contact zones in Asia pointed towards some of the same genes acting in the *Wnt* signalling component of the melanogenesis pathway³ known to modify pigmentation patterns⁹.

Yet, F_{ST} outliers do not necessarily reflect the quantitative trait loci underlying phenotypic variation¹⁰. Admixture analyses at contact zones are a powerful means to disassociate population identity from phenotype and firmly establish whether genes coding for phenotypic divergence are indeed subject to divergent selection¹¹. Here, we draw from information of admixed genomes to fine-map the genetic basis of phenotypic divergence and infer the mode and strength of selection modulating contemporary gene flow across the contact zone^{11,12}. We sampled a total of 409 individuals from replicate transects in the central ($N = 173$) and southern part of 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 1111 SNPs selected from 16.6 million variants segregating in European populations including highly ancestry-informative (high F_{ST} $N = 735$; fixed $N = 51$) and background markers ($N = 325$) spread across the genome (Supplementary Fig. 3, Supplementary Table 1). Markers were equally spread across the genome except for chromosome 18 that was densely covered with 230 tightly linked markers in the previously identified outlier region⁸ (median linkage disequilibrium (LD) central $r^2 = 0.18$, southern $r^2 = 0.17$). For a set of downstream analyses, chromosome 18 was treated separately from the remaining genome-wide SNPs (median LD within chromosomes: $r^2 = 0.0024$).

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 feathering¹³ (Supplementary Fig. 4). Inter-individual variation in shading was

highly correlated among patches (average Spearman's $r = 0.76$, Supplementary Fig. 5a). Next, we performed 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 an overall 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 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 of 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, Supplementary Table 2). High linkage disequilibrium among SNPs in this region and low levels of inter-specific recombination allowed categorizing this region by “ancestry diplotypes” (Fig. 1c): a purebred carrion crow type homozygous for the “dark” (D) allele (*chr18_{DD}*), a purebred hooded crow type homozygous for the “light” (L) allele (*chr18_{LL}*) and the heterospecific type (*chr18_{DL}*). These ancestry types predicted segregating phenotypic variation in PC1 equally well as the most significant SNPs (74.29% vs. 76.07%, respectively). *Chr18_{DD}* individuals ($N = 29$) were completely black corresponding to a narrow distribution of high PC1 scores (standard deviation [SD] in PC1 scores = 0.07, range 3.96–4.33) (Fig. 1d). On the contrary, *chr18_{LL}* individuals had a much wider phenotypic distribution (SD = 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. The upper belly and the mantle, however, were grey in all cases. Individuals of the heterospecific *chr18_{DL}* type covered almost the entire phenotypic spectrum except for the extremes of pure black or the grey-coated hooded phenotypes (SD 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 *chr18_{LL}* and *chr18_{DL}* 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 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 had been hypothesized to be involved in phenotypic divergence across multiple independent crow hybrid zones^{3,8}. Gene expression of *NDP* had 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 to modulate both the intensity and position of plumage pigmentation, consistent with the wide phenotypic distribution in *chr18_{LL}* and *chr18_{DL}* individuals (Fig. 1d).

Next, we included dominance and epistasis into the statistical model. The best model accounted for 87.91% of the variance in PC1 corresponding to $78.22\% \times 87.91\% = 68.76\%$ of the total phenotypic variance. It supported additive and dominance effects of *LRP6*, *NDP* and *chr18*, as well as recessive epistasis between *NDP* and *chr18* (Supplementary Table 2).

Allelic variation in all three SNPs linked to *NDP* had no phenotypic effect in all-black *chr18_{DD}* individuals, but accounted for most of the residual variation in *chr18_{DL}* and *chr18_{LL}* (Fig. 1d). Individuals of both diplotypes (*chr18_{DL}* and *chr18_{LL}*) became increasingly lighter in combination with the *NDP_{DL}* and *NDP_{LL}* genotype. Only *chr18_{LL}* individuals homozygous for the light *NDP* allele (i.e., *NDP_{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^{3,8}. Moreover, the 170 significant SNPs identified in the GWAS were among those having the highest F_{ST} values (GLMM: P-value $< 2 \times 10^{-16}$; Supplementary Fig. 7). F_{ST} outliers need not be associated with the genes underlying phenotypic divergence, in particular for polygenic trait architectures¹⁰. 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 upon admixed genomes segregating in the contemporary hybrid zone, we performed cline 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 cline function (Fig. 2a,e, Supplementary Table 3) supporting previous evidence for recurrent backcrossing³. Inferred cline widths of 316.43 km (central) and 553.85 km (southern) substantially exceed 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^{8,16}. In contrast to neutral genetic variation, barrier loci under divergent selection that confer a cost to hybrids would fail to move freely across hybrid zones, and are thus expected to show steeper clines and reduced width while maintaining the centre of the cline¹². 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, 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 for *chr18* (Fig. 2f, Supplementary Fig. 8, Supplementary Table 5), where the formal estimate of 241.35 km was likely inflated due to incomplete sampling of populations with carrion crow ancestry (Fig. 2f).

Genomic cline 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 admixture¹⁷. Loci subject to divergent selection are associated with lower fitness in heterospecific genomic backgrounds and are accordingly less likely to introgress into the alternative genomic background. Using the full multi-locus data set, we fitted genomic cline models that describe variation in cline shape with two parameters u and v . Analogous to a shift in cline centre and width in a geographic context, u and v describe the departure of introgression from the genome average. Elevated values of the genomic cline rate v are consistent with underdominance and epistatic incompatibilities, whereas changes in the u parameter are indicative of either directional selection or unequal

sampling¹¹. All loci that showed significantly steeper and thus narrower clines (larger v) than the genome average in the south were located in the vicinity of *chr18*, *NDP* and *LRP6*. In the central hybrid zone, patterns were qualitatively similar, but due to the smaller sample size not statistically significant (Fig. 2c,d,g,h, Supplementary Table 6). These results thus complement evidence from the geographic cline analyses supporting divergent selection on all three loci associated with phenotypic variation.

So far, we have treated *chr18* as a single locus. This region is characterized by a saddle pattern of F_{ST} peaks and contains three melanogenesis genes acting upstream of the transcription factor *MITF*, which is known to play a central role in crow phenotypic divergence^{8,9,13} (Supplementary Figs. 9–11). Pinpointing the genes responsible for the large contribution of *chr18* to phenotypic variation is impeded by high linkage disequilibrium owing to low recombination, possibly due to an inversion⁸ or proximity to the centromere¹⁸. Nevertheless, *AXIN2*, located under the first F_{ST} peak, is a prime candidate. Like *NDP*, it acts in the *Wnt* pathway and is involved in pigmentation and patterning¹⁹. *PRKCA* and a tandem array of *CACNG* genes act in the MAP-Kinase pathway and reside in the second F_{ST} peak. This region was also characterized by signatures of positive selection in the lineage leading to the hooded crow (Supplementary Figs. 9h and 10h) and shows the strongest association of all SNPs on chromosome 18 with *NDP* ($P = 2 \times 10^{-7}$). While this favours mutations in *CACNG/PRKCA* as causal in explaining phenotypic variation, a substantially larger sample of recombinant individuals is needed for fine mapping. Our results even raise the possibility that an additional epistatic interaction between *AXIN2* and *PRKCA/CACNG* might be responsible for phenotypic divergence (Supplementary Text).

This study demonstrates the power of leveraging information from naturally occurring hybridization for speciation research¹¹. In combination with outlier scans of parental populations, admixture analyses provide valuable insight into the genetic architecture of phenotypic variation, its genetic control and selection acting upon it. Variation in hybrid colour phenotypes was best attributed to recessive epistasis between two genes predicting the two pure parental forms and three hybrid classes. This corresponds well with previous, morphology-based classification into three distinct hybrid types^{20,21}. It also adds to the evidence that gene-by-gene interactions are an important component of phenotypic variation and may help explain the missing heritability for quantitative traits²². In contrast to indirect evidence from genome-scans and gene expression analysis from previous work^{3,8,9}, this study provides conclusive evidence that genes controlling phenotypic variation are subject to divergent selection. There is considerable overlap in the genes identified by genome scans comparing morphologically pure parental populations^{3,8} and cline analysis in admixed genomes (this study; Pearson's correlation between $-\log_{10}(P\text{-values})$ of v with F_{ST} central = 0.71 and south = 0.80). This supports the view that in species that have experienced recent gene exchange the much debated regions of elevated genetic differentiation (“speciation islands”²³) may constitute a valid starting point in the quest for barrier loci relevant to species diversification^{1,24}. Finally, it is striking that all genomic regions identified to be under divergent selection controlled phenotypic variation resembling findings in a pair of North American parulid warblers²⁵. Considering that divergence in plumage pigmentation patterns reduces gene flow in crows^{5,6,26–28}, this finding supports the hypothesis that

population divergence may start from very few, large-effect loci conveying prezygotic isolation²⁹. It thus strengthens the notion that strong prezygotic barriers may evolve prior to postzygotic isolation³⁰.

Methods

Population sampling

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, $N = 30$). For all these individuals, blood samples were taken from the brachial vein and stored either in Queen's lysis buffer, EDTA- 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 16 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.

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 Beringungserl-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 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.

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 was assessed using Nanodrop and the SYBR green fluorescence assay (Invitrogen).

From a total of 16.6 million variants segregating in the European crow populations^{3,8} we selected 1152 SNPs for genotyping with the GoldenGate assay (Illumina). Golden gate primers were primarily designed to contain information on ancestry differentiating between carrion and hooded crows. Using prior knowledge on the degree of differentiation (F_{ST}) per SNP we primarily chose SNPs with high allelic differential, but also included a set of ancestry-informative makers representing the genomic background (Supplementary Fig. 3). Specifically, the following types of makers were chosen: *i) fixed* ($N=51$) — variants that were fixed between European carrion and hooded crows⁸. *ii) outlier* ($N=767$) — a taxon-informative SNP set differentiating between carrion and hooded crows without fixed differences. This set included SNPs with the highest F_{ST} values (99th percentile) residing within localized phylogenetic trees (class I “cacti”) or 50 kb outlier windows with elevated genetic differentiation (95th percentile of windows-based F_{ST} ; see 3,8). This set also included 4 SNPs within *candidate* melanogenesis genes that were not part of class I cacti or outlier windows. *iii) background* ($N=334$) — a genome-wide SNP-set that has not been shown to be under selection⁸, but that was still taxon-informative, was defined as variants that fell within the 80–95th F_{ST} percentile in regions 100 kb away from outlier regions and not within 10 kb of each other. These included another 14 SNPs residing within *candidate* melanogenesis genes. Note that due to an extremely right skewed F_{ST} distribution, the F_{ST} values of the 80–95th percentile are low and reflective of the genome-wide average (for F_{ST} values see below).

In a first round of filtering, we excluded SNPs with a minor allele frequency < 0.1 and SNPs within 100 bp of each other. Forward and reverse primers spanning 60 bp around the target loci were designed with Primer3³¹. In a second filtering step, we excluded SNPs where the primers included blocks of missing sequences or did not map uniquely in the genome. We further removed SNPs with a GoldenGate primer feasibility score < 0.6 , which also excluded failed primer designs (Supplementary Table 1). For the *background* variant set, we chose only one SNP per gene. For the *fixed* and *outlier* markers and *candidate* genes, we allowed one variant with the highest primer design score in each gene region (e.g. exon, intron, regulatory region defined as 5 kb surrounding a gene), which resulted in a higher density within the peak region on chromosome 18.

The exact location of each SNP in reference to genome version 2.5 (RefSeq Assembly ID on NCBI: GCF_000738735.1 as published in 8) 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 (v2011.1). We removed 31 individuals that had a call rate of 0%, and two 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, F_{ST} ranged from 0.766–1.000 (mean = 0.907) and 0.844–1.000 (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 2217 comparisons suggesting an accuracy of genotype calls above 99.95%. Of the allopatric populations, 40 individuals had been previously genotyped using the HaplotypeCaller in GATK (v3.3.032) after paired-end whole-genome sequencing on the HiSeq2000 (Illumina) platform³. Coverage ranged from 6.83× to 23.45× (average = 11.37×, median = 10.64×). Five of those individuals were additionally genotyped using GoldenGate. In those, we found 104 inconsistencies out of 5501 genotypes (1.89%). For our final data set we kept the genotype calls from GoldenGate wherever possible, such that we had $N = 522$ individuals genotyped at $N = 1111$ polymorphic loci (average call rate of 99.32 %). The distribution of the 1111 SNPs in the genome is illustrated in Supplementary Fig. 3. Genotypic data is available in Supplementary Table 8.

Analyses

Hybrid indices and hybrid classes—We estimated maximum likelihood hybrid indices separately for the central and southern hybrid zone including all nest mates. Allopatric German, Polish and Swedish samples were used as ancestral reference populations in the central zone, Italian and different allopatric German populations for the southern zone, respectively. Previous work has shown that most of the genetic variation separating carrion and hooded crows clusters within a ~2 Mb region on chromosome 188. Therefore, we estimated hybrid indices separately for SNPs on chromosome 18, containing the most ancestry informative SNPs, and the rest of the autosomal SNPs with the R-package *introgress* (v1.2.333) in R (v3.4.334). In the following, we refer to these hybrid indices as *chr18* and *GW* (genome wide), respectively. A hybrid index of 0 represents a purebred carrion crow and a hybrid index of 1 a purebred hooded crow.

Using the *NewHybrids* software (v2.0+ Developmental, July/August 2007) we further estimated the posterior probability of an individual belonging to one out of six distinct genealogical classes (purebred carrion or hooded crows, F1- or F2-hybrids, and backcrosses to carrion or hooded crows) by simulating the different genotype frequency classes³⁵. We used the *parallelnewhybrid* (v0.0.0.900236) R-package to call *NewHybrids* with uninformative Jeffreys-type priors for the estimation of allele frequencies (Θ) and mixing proportions of genotype frequency classes (π). We ran *NewHybrids* separately for the central and southern hybrid zone and separately for SNPs on chromosome 18 and the rest of the autosomal SNPs, discarding the first 20,000 generations as burn-in and estimating parameters from the following 200,000 MCMC algorithm iterations. Assignment efficiency, accuracy and overall performance were assessed by simulating multigenerational hybrid data sets³⁷. For that we used the *hybriddetective* (v0.1.0.900038) R-package implementing the *HybridLab* algorithm³⁹. Using the genotype data of the allopatric samples (see above), we simulated four times $N = 500$ purebred (that is carrion or hooded crows), $N = 250$ F1, $N = 125$ F2 and $N = 125$ BC individuals. The relative frequencies of the different genealogical classes were matched to those inferred empirically by *NewHybrids*, the performance of

which is known to be affected by the proportion of hybrids in the sample³⁷. Following the procedure used for the empirical data, we ran all simulations separately for the central and southern hybrid zone and separately for SNPs on chromosome 18 and the rest of the autosomal SNPs. We then combined these simulated data sets with the allopatric samples and ran NewHybrids using the same settings as for the empirical data. We merged the results of the four simulated data sets ($4 \times$ central autosomal SNPs, $4 \times$ southern autosomal SNPs, $4 \times$ central chromosome 18 SNPs, $4 \times$ 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 0.995 across all genealogical classes for both SNP sets in the two hybrid zones. Using this cut-off, NewHybrids assigned 99.14% of purebred, 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 heterozygous at the outlier loci on chromosome 18. Using the SNPs on chromosome 18, NewHybrids assigned all individuals into 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, removing nest mates; Supplementary Table 9). We call these clear diplotypes based on chromosome 18 *chr18_{DD}*, *chr18_{LL}* and *chr18_{DL}*, respectively.

We additionally used these plots to identify inter-specific 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 tracts of at least three SNPs, single SNP changes were discarded as genotype errors or variation. Assuming randomness of cross-over locations, 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 (GWA)—Digital photographs of the ventral and dorsal side were taken for all birds from Graz and for 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 amount of grey and black in the feathering (see below). For each of the 129 individuals we scored seven plumage patches on the ventral side and four regions on the dorsal side for the amount of grey and black in the feathering (Fig. 1a, 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 `missMDA` R-package (v1.1140). Individual measurements were then summarized using a principal component analysis (PCA) as implemented in the `FactoMineR` R-package (v1.3941). All

eleven variables loaded positively on principal component 1 (PC1), which explained 78.22% of the variance in pigmentation pattern and intensity. PC2 explained 7.66% of the variance and separated the four belly quadrants and the mantle from the other plumage patches. In order to assess the objectivity of scoring, we estimated inter-observer repeatability of the colour PC1 and PC2 scores by measuring a subset of individuals by a second person (N.S., $N=105$ individuals). We estimated the repeatability of PC1 and PC2 scores using the `rptGaussian()` function from the `rptR` R-package (v0.9.2142). Both PC1 and PC2 scores were highly repeatable between observers (PC1: $r \pm SE = 0.965 \pm 0.007$, $P = 4 \times 10^{-64}$; PC2: $r \pm SE = 0.646 \pm 0.057$, $P = 1 \times 10^{-14}$ using $N=10,000$ parametric bootstraps for interval 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 one degree of freedom (additive effect). Models were fitted with the `qtscore()` function and a Gaussian error structure in the `GenABEL` R-package (v1.8-043). Since SNPs showed varying degrees of linkage disequilibrium, we used the `simpleM`-algorithm with default settings to estimate the effective number of tests performed⁴⁴ and used this estimate for controlling the genome-wide type I error rate.

We then fitted all SNPs showing a significant additive main effect on colour PC1 (3 SNPs on chromosome 1, 1 SNP on chromosome 1A, 166 SNPs on chromosome 18) as factors to test for additional dominant gene action and tested epistatic interactions between all of them ($3 \times 1 + 3 \times 166 + 1 \times 166 = 667$ interactions). For significance testing, we performed likelihood ratio tests comparing models with and without the dominant gene action or interaction. Because both chromosome 18 and the three SNPs on chromosome 1 effectively behave 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 $AIC - 2$ as selection threshold⁴⁵.

Under Hardy-Weinberg-equilibrium the variance explained by a SNP is calculated as $V_{SNP} = 2 \times p \times (1 - p) \times \beta^2$, where p is the minor allele frequency and β the estimated slope for the SNP-effect (i.e. the average effect of an allelic substitution⁴⁶) from the above regression models. To obtain the proportion of phenotypic variance explained by a SNP, V_{SNP} is divided by the phenotypic variance, which is equivalent to the multiple R-squared (r^2) from an ordinary least squares regression model. In our study populations, however, most SNPs were not in Hardy-Weinberg-equilibrium and we thus used the multiple R-squared (r^2) from the regression models to get an estimate of the variance explained by additive, dominance or interaction effects.

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 GWAS 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 principal component analyses for both the central and southern hybrid zone separately (without the allopatric populations) and reconstructed coordinates using PC1 only. We then calculated great-circle (orthodromic) distances in km

between these reconstructed coordinates. This essentially corresponds to a perpendicular projection of sampling locations onto a transect minimizing the distance to the sampling locations. We added the allopatric populations by calculating their great-circle distances from the respective ends of the sampling transect. We then fitted geographic cline models using the *hzar* (v0.2-547) package in R using the mean *chr18* and *GW* hybrid indices per sampling location as the dependent variable and geographic distance between sampling locations (in km) as the predictor. We included only one individual per nest in order to reduce pseudoreplication (i.e. we kept $N = 116$ individuals in the central and $N = 273$ individuals in the southern hybrid zone). In *hzar*, cline models as described in 17,48 are fitted using the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm^{49,50}. Model selection between the null model and three cline models (I–III) varying in the number parameters was based on Akaike's information criterion using $AIC - 2$ as selection threshold⁴⁵. In model I, we fixed the minimum and maximum hybrid index frequencies p_{\min} and p_{\max} to their minimum and maximum observed mean values (scaling = fixed) and fitted a model without exponential tails (tails = none). Thus, in model I only the sigmoid cline function is fitted and the two parameters cline centre and cline width are estimated. In model II we additionally estimated the minimum and maximum hybrid index frequencies p_{\min} and p_{\max} (scaling = free) but did not include exponential tails (tails = none). In model III we also estimated the minimum and maximum hybrid index frequencies p_{\min} and p_{\max} and fitted exponential tails with independent size (δ_{left} , δ_{right}) and shape (τ_{left} , τ_{right}) parameters (tails = both). In one case (southern hybrid zone, mean *chr18* hybrid index) AIC was 1.89 and we present results of model I which had the non-significant better fit.

Genomic clines—Genomic cline models use the hybrid index instead of the geographic distance to detect loci that may be subject to selection. We fitted genomic cline models both with the *gghybrid* R-package (v0.0.0.900051) and with *bgc* (v1.0352) for both the central and the southern hybrid zone. *gghybrid* allows fitting hybrid indices that have been estimated using only a subset of the markers. Previous work suggests that markers in the outlier region on chromosome 18 are in significant linkage disequilibrium and putatively under positive selection⁸. To test against a genome-wide background of neutral genetic variation we fitted models in *gghybrid* with hybrid indices estimated from all SNPs excluding chromosome 18 (but including chromosome Z). *gghybrid* uses the same method for estimating hybrid indices as the *introgress* R-package (with an additional prior⁵³) and hybrid indices were highly correlated between the two packages ($r = 0.99$). We then fitted genomic cline models including only a single individual per nest and the allopatric populations, discarding the first 10,000 iterations as burn-in and estimating parameters and posterior probabilities (P-values) from the following 40,000 MCMC iterations. *gghybrid* fits Fitzpatrick's logit-logistic cline function⁵⁴ and estimates parameters v and u . Parameter v is always positive and higher values indicate steeper clines. Parameter u is related to the centre of the cline, i.e. the hybrid index at which the allele frequencies are half way between those of the allopatric parental populations.

The Bayesian genomic cline model implemented in *bgc* derives hybrid indices internally using all markers and then fits the Barton cline model¹⁷. Thus, the *bgc* hybrid indices integrate over both the hybrid indices estimated from SNPs on chromosome 18 and those

estimated using all remaining autosomal SNPs. *bgc* derives parameters and no P-values and parameter estimates should be unaffected by related individuals. Thus, we included all nest mates in these models. *bgc* models were fitted with the ICARrho model for linked loci, discarding the first 250,000 generations as burn-in, estimating parameters from the following 500,000 MCMC algorithm iterations and recoding every 100th value and default parameters otherwise. We ran two independent MCMC simulations per hybrid zone and combined the output of the two chains. We estimated means and 95% confidence intervals for the slope parameter β , which describes the rate at which the probability of ancestry transitions. Positive values indicate steeper clines and a reduced rate of introgression⁵⁵. Since we used all SNPs including those under divergent selection on chromosome 18, β is downward biased due to the fact that SNPs are densely spaced on chromosome 18, show the strongest clinal variation and contribute disproportionately to the hybrid index (correlation of internally derived hybrid index with *chr18*: $r = 0.94$ and *GW*: $r = 0.84$ for the central hybrid zone and *chr18*: $r = 0.90$ and *GW*: $r = 0.82$ for the southern hybrid zone). For visual representation we thus normalized the distribution of β -estimates by the median. Outliers were classified as loci falling outside the empirical 95% quantile distribution. The analysis was repeated excluding SNPs on chromosome 18, in which case the correlation of internally derived hybrid indices with *GW* was much higher ($r = 0.97$ for the central hybrid zone and $r = 0.96$ for the southern hybrid zone).

Population genetic analyses

Ancestral state reconstruction: We determined the ancestral state of all SNPs in the GoldenGate assay using the data and same methods as in 8 for five genomes of the American crow (*Corvus brachyrhynchos*), two rook genomes (*Corvus frugilegus*) and two genomes of the jackdaw (*Corvus monedula*). 80.72% of the sites were fixed in all three outgroup species (73.29% in 8) and thus informative for the ancestral state in hooded and carrion crows. 11.84% of the sites were polymorphic in American crows but fixed for the same allele in rooks and jackdaws, which most likely represent ancestral polymorphisms prior to the divergence of American crows from hooded and carrion crows. These sites are thus uninformative in respect to the ancestral state in hooded and carrion crows. 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 (homoplasy) 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 data³ using PLINK (v1.90b4.456). We used the allopatric samples of carrion and hooded crows from the central and southern hybrid zone independently as input populations. F_{IS} values were estimated from chromosome 18 genotypes of individuals sampled in the central and southern hybrid zone and that were assigned either a hooded or carrion crow or F1-hybrid ancestry. For that we used the R-package hierfstat (v0.04-2257).

Principal component analyses (PCA): To better understand the partitioning of genetic variation along chromosome 18 (see 58,59), we performed PCA using all individuals from the two hybrid zones ($N=173$ individuals and $N=236$ individuals for the central and southern zone, respectively) and all SNPs on chromosome 18 for which we could resolve the ancestral state and that had a missing call rate smaller 0.05 and a minor allele frequency larger than 0.05 ($N=204$ SNPs and $N=193$ SNPs for the central and southern zone, respectively). Discrete clusters of genetic variation provide evidence for divergent haplotypes that are non-recombining (for example due to an inversion). In case of two major haplotypes, we expected principal component 1 (PC1) to classify individuals by diplotype: homozygous individuals are expected to cluster at both ends of the PC1 distribution with heterozygous individuals in between. We further expect the squared principal component loadings of PC1 to reflect F_{ST} (termed communality h^2 ; 60), 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. By using the ancestry information, the principal component loadings can be polarized. 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 (v1.12.161).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study would not have been possible without the commitment of dedicated ornithologists who helped locating active nest sites and participated in sampling. These include Manfred Hug and colleagues in Brandenburg, Jens Voigt, Dieter Kronbach, Jörg Wollmerstädt, Matthias Schrack and Winfried Nachtigall in Sachsen, Markus Döpfner in Baden-Württemberg, Sebastian Zinko and Monika Grossmann in Austria and Amministrazione Provinciale of Alessandria, Asti and Cuneo in Italy. We would further like to acknowledge the Max Planck Institute for Ornithology in Radolfzell, the Friedrich-Löffler-Institute and the Förderverein Sächsische Vogelschutzwerke 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 the Swedish National Infrastructure for Computing provided access to computational resources. Funding was provided by the Volkswagen Stiftung (grant I/83 496 to JW), the European Research Council (ERCStG-336536 FuncSpecGen to JW), the Knut and Alice Wallenberg Foundation (project grant including JW) and LMU Munich (to JW).

References

1. Wolf JBW, Ellegren H. Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet.* 2017; 18:87–100. DOI: 10.1038/nrg.2016.133 [PubMed: 27840429]
2. Ravinet M, et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol.* 2017; 30:1450–1477. DOI: 10.1111/jeb.13047 [PubMed: 28786193]
3. Vijay N, et al. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat Commun.* 2016; 7:e13195.doi: 10.1038/ncomms13195
4. Mayr, E. Systematics and the origin of species. Columbia University Press; 1942.
5. Randler C. Assortative mating of carrion *Corvus corone* and hooded crows *C. cornix* in the hybrid zone in eastern Germany. *Ardea.* 2007; 95:143–149. DOI: 10.5253/078.095.0116
6. Saino N, Scatizzi L. Selective aggressiveness and dominance among carrion crows, hooded crows and hybrids. *B Zool.* 1991; 58:255–260. DOI: 10.1080/11250009109355762

7. Londei T. Alternation of clear-cut colour patterns in *Corvus* crow evolution accords with learning-dependent social selection against unusual-looking conspecifics. *Ibis*. 2013; 155:632–634. DOI: 10.1111/ibi.12074
8. Poelstra JW, et al. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*. 2014; 344:1410–1414. DOI: 10.1126/science.1253226 [PubMed: 24948738]
9. Poelstra JW, Vijay N, Hoepfner MP, Wolf JBW. Transcriptomics of colour patterning and coloration shifts in crows. *Mol Ecol*. 2015; 24:4617–4628. DOI: 10.1111/mec.13353 [PubMed: 26302355]
10. Le Corre V, Kremer A. The genetic differentiation at quantitative trait loci under local adaptation. *Mol Ecol*. 2012; 21:1548–1566. DOI: 10.1111/j.1365-294x.2012.05479.x [PubMed: 22332667]
11. Gompert Z, Mandeville EG, Buerkle CA. Analysis of population genomic data from hybrid zones. *Annu Rev Ecol Evol Syst*. 2017; 48:207–229. DOI: 10.1146/annurev-ecolsys-110316-022652
12. Barton NH, Hewitt GM. Analysis of hybrid zones. *Annu Rev Ecol Syst*. 1985; 16:113–148. DOI: 10.1146/annurev.es.16.110185.000553
13. Wu C-C, et al. In situ quantification of individual mRNA transcripts in melanocytes discloses gene regulation of relevance to speciation. *J Exp. Biol*. 2019; doi: 10.1242/jeb.194431
14. Ke JY, et al. Structure and function of Norrin in assembly and activation of a Frizzled 4–Lrp5/6 complex. *Gene Dev*. 2013; 27:2305–2319. DOI: 10.1101/gad.228544.113 [PubMed: 24186977]
15. Vickrey AI, et al. Introgression of regulatory alleles and a missense coding mutation drive plumage pattern diversity in the rock pigeon. *Elife*. 2018; 7:e34803.doi: 10.7554/eLife.34803 [PubMed: 30014848]
16. Haas F, et al. An analysis of population genetic differentiation and genotype-phenotype association across the hybrid zone of carrion and hooded crows using microsatellites and *MC1R*. *Mol Ecol*. 2009; 18:294–305. DOI: 10.1111/j.1365-294x.2008.04017.x [PubMed: 19076276]
17. Szymura JM, Barton NH. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina orientalis* and *B. orientalis*, near Cracow in southern Poland. *Evolution*. 1986; 40:1141–1159. DOI: 10.2307/2408943 [PubMed: 28563502]
18. Weissensteiner MH, et al. Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. *Genome Res*. 2017; 27:697–708. DOI: 10.1101/gr.215095.116 [PubMed: 28360231]
19. Kofron M, et al. The role of maternal axin in patterning the *Xenopus* embryo. *Dev Biol*. 2001; 237:183–201. DOI: 10.1006/dbio.2001.0371 [PubMed: 11518515]
20. Duquet M. Pièges de l'identification: La Corneille mantelée *Corvus cornix*: pure ou hybride? *Ornithos*. 2012; 19:57–67.
21. Rolando A. A study on the hybridization between carrion and hooded crow in northwestern Italy. *Ornis Scand*. 1993; 24:80–83. DOI: 10.2307/3676414
22. Mackay TFC. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet*. 2014; 15:22–33. DOI: 10.1038/nrg3627 [PubMed: 24296533]
23. Pennisi E. Disputed islands. *Science*. 2014; 345:611–613. DOI: 10.1126/science.345.6197.611 [PubMed: 25104366]
24. Nadeau NJ, et al. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Res*. 2014; 24:1316–1333. DOI: 10.1101/gr.169292.113 [PubMed: 24823669]
25. Toews DPL, et al. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Curr Biol*. 2016; 26:2313–2318. DOI: 10.1016/j.cub.2016.06.034 [PubMed: 27546575]
26. Brodin A, Haas F. Speciation by perception. *Anim Behav*. 2006; 72:139–146. DOI: 10.1016/j.anbehav.2005.10.011
27. Brodin A, Haas F. Hybrid zone maintenance by non-adaptive mate choice. *Evol Ecol*. 2009; 23:17–29. DOI: 10.1007/s10682-007-9173-9
28. Saino N, Villa S. Pair composition and reproductive success across a hybrid zone of carrion crows and hooded crows. *Auk*. 1992; 109:543–555.
29. Seehausen O, et al. Speciation through sensory drive in cichlid fish. *Nature*. 2008; 455:620–626. DOI: 10.1038/nature07285 [PubMed: 18833272]

30. Coyne JA, Orr HA. “Patterns of speciation in *Drosophila*” revisited. *Evolution*. 1997; 51:295–303. DOI: 10.2307/2410984 [PubMed: 28568795]
31. Untergasser A, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res*. 2012; 40:e115.doi: 10.1093/nar/gks596 [PubMed: 22730293]
32. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011; 43:491–498. DOI: 10.1038/ng.806 [PubMed: 21478889]
33. Gompert Z, Buerkle CA. introgress: a software package for mapping components of isolation in hybrids. *Mol Ecol Resour*. 2010; 10:378–384. DOI: 10.1111/j.1755-0998.2009.02733.x [PubMed: 21565033]
34. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2017.
35. Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*. 2002; 160:1217–1229. [PubMed: 11901135]
36. Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR. *parallelnewhybrid*: an R package for the parallelization of hybrid detection using NewHybrids. *Mol Ecol Resour*. 2017; 17:91–95. DOI: 10.1111/1755-0998.12597 [PubMed: 27617417]
37. Vähä JP, Primmer CR. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Mol Ecol*. 2006; 15:63–72. DOI: 10.1111/j.1365-294x.2005.02773.x [PubMed: 16367830]
38. Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR. *hybriddetective*: a workflow and package to facilitate the detection of hybridization using genomic data in R. *Mol Ecol Resour*. 2017; 17:e275–e284. DOI: 10.1111/1755-0998.12704 [PubMed: 28776912]
39. Nielsen EE, Bach LA, Kotlicki P. HybridLab (version 1.0): a program for generating simulated hybrids from population samples. *Mol Ecol Notes*. 2006; 6:971–973. DOI: 10.1111/j.1471-8286.2006.01433.x
40. Josse J, Husson F. missMDA: a package for handling missing values in multivariate data analysis. *J Stat Softw*. 2016; 70:1–31. DOI: 10.18637/jss.v070.i01
41. Lê S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. *J Stat Softw*. 2008; 25:1–18. DOI: 10.18637/jss.v025.i01
42. Stoffel MA, Nakagawa S, Schielzeth H. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol Evol*. 2017; 8:1639–1644. DOI: 10.1111/2041-210x.12797
43. Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007; 23:1294–1296. DOI: 10.1093/bioinformatics/btm108 [PubMed: 17384015]
44. Gao XY, Stamier J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol*. 2008; 32:361–369. DOI: 10.1002/gepi.20310 [PubMed: 18271029]
45. Burnham, KP, Anderson, DR. Model selection and multimodel inference. 2nd edn. Springer; 1998.
46. Lynch, M, Walsh, B. Genetics and analysis of quantitative traits. Sinauer; 1998.
47. Derryberry EP, Derryberry GE, Maley JM, Brumfield RT. hzar: hybrid zone analysis using an R software package. *Mol Ecol Resour*. 2014; 14:652–663. DOI: 10.1111/1755-0998.12209 [PubMed: 24373504]
48. Szymura JM, Barton NH. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution*. 1991; 45:237–261. DOI: 10.1111/j.1558-5646.1991.tb04400.x [PubMed: 28567861]
49. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys*. 1953; 21:1087–1092. DOI: 10.1063/1.1699114
50. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. 1970; 57:97–109. DOI: 10.2307/2334940
51. Bailey R. gghybrid: evolutionary analysis of hybrids and hybrid zones. R package version 0.0.0.9000. 2018

52. Gompert Z, Buerkle CA. bgc: software for Bayesian estimation of genomic clines. *Mol Ecol Resour.* 2012; 12:1168–1176. DOI: 10.1111/1755-0998.12009.x [PubMed: 22978657]
53. Buerkle CA. Maximum-likelihood estimation of a hybrid index based on molecular markers. *Mol Ecol Notes.* 2005; 5:684–687. DOI: 10.1111/j.1471-8286.2005.01011.x
54. Fitzpatrick BM. Alternative forms for genomic clines. *Ecol Evol.* 2013; 3:1951–1966. DOI: 10.1002/ece3.609 [PubMed: 23919142]
55. Sung CJ, Bell KL, Nice CC, Martin NH. Integrating Bayesian genomic cline analyses and association mapping of morphological and ecological traits to dissect reproductive isolation and introgression in a Louisiana Iris hybrid zone. *Mol Ecol.* 2018; 27:959–978. DOI: 10.1111/mec.14481 [PubMed: 29319908]
56. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575. DOI: 10.1086/519795 [PubMed: 17701901]
57. Goude J, Jombart T. hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.04-22. 2015
58. Ma JZ, Amos CI. Investigation of inversion polymorphisms in the human genome using principal components analysis. *Plos One.* 2012; 7:e40224.doi: 10.1371/journal.pone.0040224 [PubMed: 22808122]
59. Knief U, et al. Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome Biol.* 2016; 17:e199.doi: 10.1186/s13059-016-1056-3
60. Duforet-Frebourg N, Luu K, Laval G, Bazin E, Blum MGB. Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 genomes data. *Mol Biol Evol.* 2016; 33:1082–1093. DOI: 10.1093/molbev/msv334 [PubMed: 26715629]
61. Zheng XW, et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics.* 2012; 28:3326–3328. DOI: 10.1093/bioinformatics/bts606 [PubMed: 23060615]

One sentence summary

Admixture mapping identifies epistatic interaction of major-effect pigmentation loci as the molecular genetic basis of prezygotic isolation.

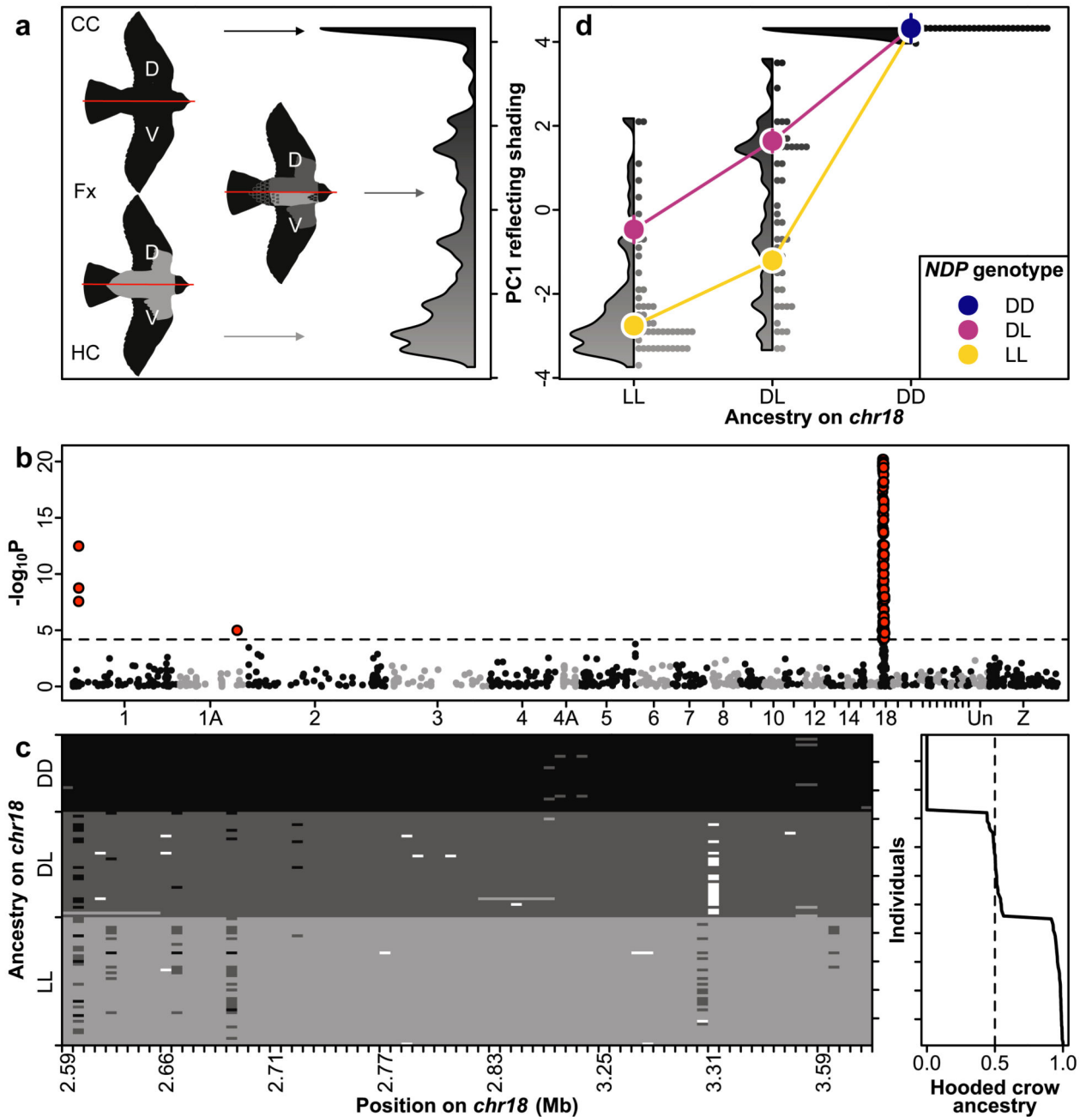


Figure 1. Genetic basis of phenotypic variation.

(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 principal component analysis. The abscissa shows the smoothed density distribution of PC1 scores for 122 individuals with the underlying shading for orientation. (b) Manhattan plot of genome-wide association analysis for PC1 scores shown for 1103 SNPs by chromosome. 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 the F_{ST} outlier region of chromosome 18 (*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 SNPs group individuals into three discrete classes corresponding to the three diplotypes: *chr18_{DD}*, *chr18_{DL}*, *chr18_{LL}*. **(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).

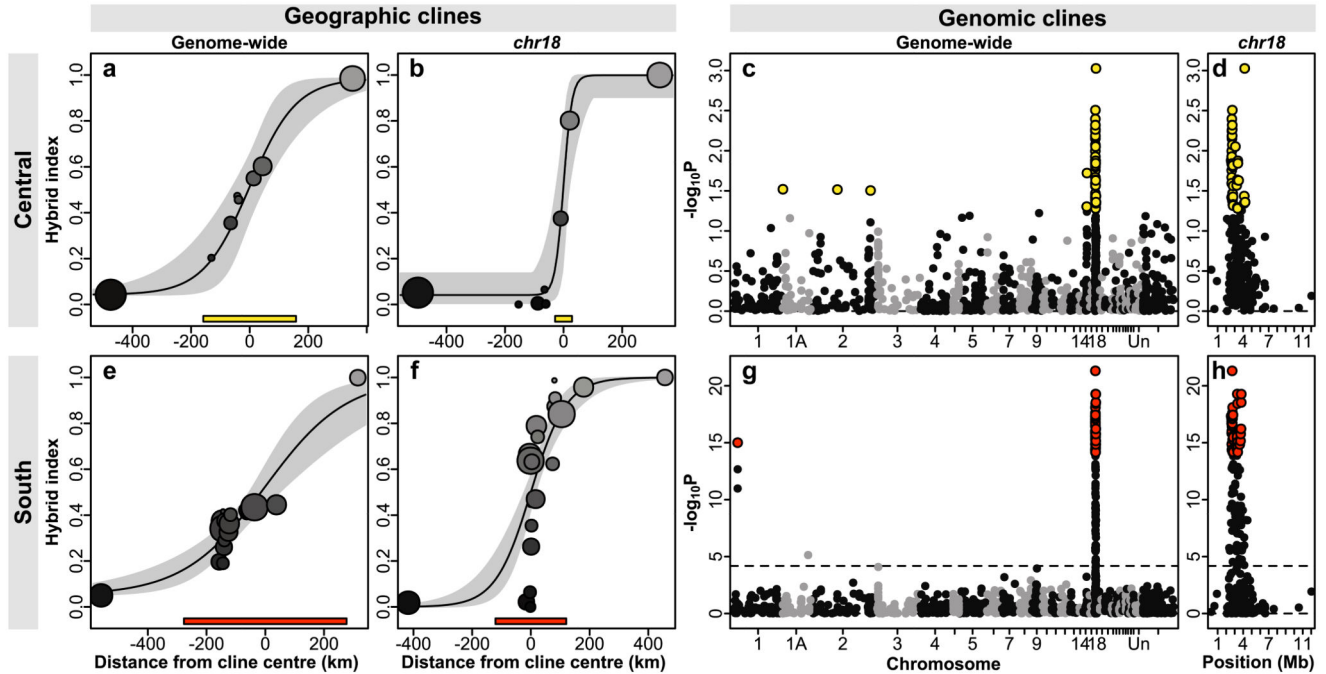


Figure 2. Cline analysis.

Left: Geographic clines for hybrid indices estimated from genome-wide data (a, e, $N = 752$ SNPs) and SNPs on chromosome 18 (b, f, $N = 230$ SNPs). Estimates are shown for both the central hybrid zone (a, b, $N = 116$ individuals) and the southern hybrid zone (e, 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 centered on the cline centre, rectangles (yellow: central, red: south) depict the cline widths predicted by the models. *Right: Summary statistics of genomic cline* analyses shown for all 1111 SNPs across the genome (left c, g) and for the outlier region of chromosome 18 (*chr18*) specifically (right d, 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.