

RESEARCH

Open Access



Human impact on the recent population history of the elusive European wildcat inferred from whole genome data

María Esther Nieto-Blázquez^{1*}, Dennis Schreiber¹, Sarah A. Mueller^{2,3}, Katrin Koch⁴, Carsten Nowak^{3,5} and Markus Pfenninger^{1,5,6}

Abstract

Background: The extent and impact of evolutionary change occurring in natural populations in response to rapid anthropogenic impact is still poorly understood on the genome-wide level. Here, we explore the genetic structure, demographic history, population differentiation, and domestic introgression based on whole genome data of the endangered European wildcat in Germany, to assess potential genomic consequences of the species' recent spread across human-dominated cultural landscapes.

Results: Reconstruction of demographic history and introgression rates based on 47 wildcat and 37 domestic cat genomes suggested late introgression between wild and domestic cat, coinciding with the introduction of domestic cat during the Roman period, but overall relatively low rates of hybridization and introgression from domestic cats. Main population divergence found between an eastern and central German wildcat clade was found to be of rather recent origin (200 y), and thus the likely consequence of anthropogenic persecution and resulting isolation in population refugia. We found similar effective population sizes and no substantial inbreeding across populations. Interestingly, highly differentiated genes between wild cat populations involved in the tryptophan-kynurenine-serotonin pathway were revealed, which plays a role in behavioral processes such as stress susceptibility and tolerance, suggesting that differential selection acted in the populations.

Conclusions: We found strong evidence for substantial recent anthropogenic impact on the genetic structure of European wildcats, including recent persecution-driven population divergence, as well as potential adaptation to human-dominated environments. In contrast, the relatively low levels of domestic introgression and inbreeding found in this study indicate a substantial level of "resistance" of this elusive species towards major anthropogenic impacts, such as the omnipresence of domestic cats as well as substantial habitat fragmentation. While those findings have strong implications for ongoing conservation strategies, we demand closer inspection of selective pressures acting on this and other wildlife species in anthropogenic environments.

Keywords: *Felis silvestris*, Whole-genome sequencing, Anthropogenic disturbance, Population divergence, Adaptive selection, Introgression, Conservation genomics

*Correspondence: menietoblazquez@gmail.com

¹ Molecular Ecology Group, Senckenberg Biodiversity and Climate Research Centre (SBIK-F), 60325 Frankfurt am Main, Germany
Full list of author information is available at the end of the article

Background

The anthropogenic alteration of ecosystems has been acknowledged as the single greatest threat to global biodiversity [1, 2], leaving a footprint at an unprecedented rate [3]. Human impact is also a well-recognized evolutionary force [4, 5], and plenty of evidence is now



available on how human disturbances alter selective pressures on populations [6–8], which results in greater and more rapid phenotypic selection than would happen in natural populations [6, 9, 10]. Threats to wildlife populations often act in a synergistic way, and genetic factors are fundamental to understand anthropogenic impact on wildlife populations and assist in the efforts to mitigate the challenges wildlife might confront [11]. As seen for several wildlife species (e.g. soay sheep, gray wolves or wildcat), introgression with domestic congeners is an indirect form of human-mediated evolution, which can pose a major conservation threat for wildlife, and indeed led to the extinction of many plants and animals [12, 13]. Despite of this human impact, many species benefit from human-modified environments, such as synanthropic species (e.g. masked palm civet, raccoon dog or red foxes [14], which show phenotypical plasticity and increased population densities [15]. In addition, several historically persecuted wildlife species of large (e.g., brown bear, gray wolf, Eurasian lynx) and medium-sized (e.g., Eurasian otter, European wildcat) carnivores currently increase in range and abundance throughout several regions, including a successful ongoing spread within the densely populated, anthropogenically transformed landscapes of Western and Central Europe.

The European wildcats (*Felis silvestris*, hereafter *FS*), for instance, suffered considerable range declines and population reductions in Central Europe, due to anthropogenic persecution as well as habitat loss and fragmentation during the 19th and early twentieth century [16, 17]. Recent studies show that despite the fragmented and mosaic-like German landscape, the species has recently reemerged in various regions [18, 19], and range expansion is still ongoing and even comprises areas with no historic presence data of *F. silvestris* [20].

Previous studies using traditional marker systems (i.e., mitochondrial control region sequences and microsatellites; [21–23]) show that wildcats in Germany are currently clustered into two main groups, a western and central one. However, little is known about the historical processes that have led to the current allopatric distribution, with no obvious landscape barrier separating both populations [18, 24]. European wildcats survived Pleistocene glaciations in three main Mediterranean refugia in southern Iberian, Italy, and Balkan peninsulas, followed by postglacial northward recolonizations [21]. This study concluded that the two populations observed in Germany are most probably due to recolonization events from different refugia. Evidence, however, suggests a recent origin of population divergence, perhaps as a consequence of massive persecution and re-expansion from different historic refugia [20].

While domestic cats (*Felis catus*, hereafter *FC*) do not directly derive from European wildcats (*FS*), successful hybridization between *FS* and *FC* has been found in various regions of their distribution. Particularly high levels of hybridization have been revealed in Scotland and Hungary [21], whereas rather low degrees of admixture were found in other European regions [21–23]. The different admixture level amongst regions can be explained by variation in population histories [21, 25], and the disparity in ecological barriers and environmental conditions [26]. However, despite of low observed recent hybridization in certain European regions, the potential effects of long-term introgression of domestic cat DNA into the gene pool of *F. silvestris* is still largely unknown. Archaeozoological data suggests the presence of domestic cat in central Europe at least during the Roman Period, ca. 2,000 ya, introduced by either Celts or Romans [27], which expanded extensively during Medieval times as a critical aid against vermin and pests [28] and was included in trades for its pelt [29]. Thus, long-term impact of domestic cat presence, in particular within the Roman-occupied areas in Germany, might have resulted in the accumulation of domestic cat ancestry within wildcat genomes, and potentially even contribute to the observed population divergence.

Here we apply a population genomic approach based on whole genome data from wildcat samples across Germany to reconstruct recent population history and assess anthropogenic impacts on the genetic integrity of the species. With the advent of high-throughput sequencing methods, conservation genomics has the potential to assist with identifying historical *versus* contemporary scenarios of population differentiation for non-model organisms [30] and also patterns of hybridization and introgression [31]. The viability of local populations of conservation concern that are impacted by climatic change and anthropogenic activities and their ability to adapt to such challenges can be predicted by identifying regions of the genomes under selection [32]. In this study we focus on i) the temporal pattern of differentiation among wildcat populations by historic (e.g. introduction of domestic cat by romans, spread of domestic cat during medieval age or Pleistocene glacial refugia) rather than contemporary events (e.g. anthropogenic pressures), ii) the selective pressures that may be associated to human influence, iii) the current level of genome-wide genetic diversity and inbreeding across the species range in Germany, which may reflect historic and recent anthropogenic impact, and iv) the proportion of domestic cat introgression and the potential role of introgression in the divergence of the two German wildcat groups. Answering those questions is of pivotal importance to understand how populations adapt and survive in human

landscapes, which may have considerable implications for conservation strategies in humans-impacted areas.

Results

Whole-genome sequencing and population structure

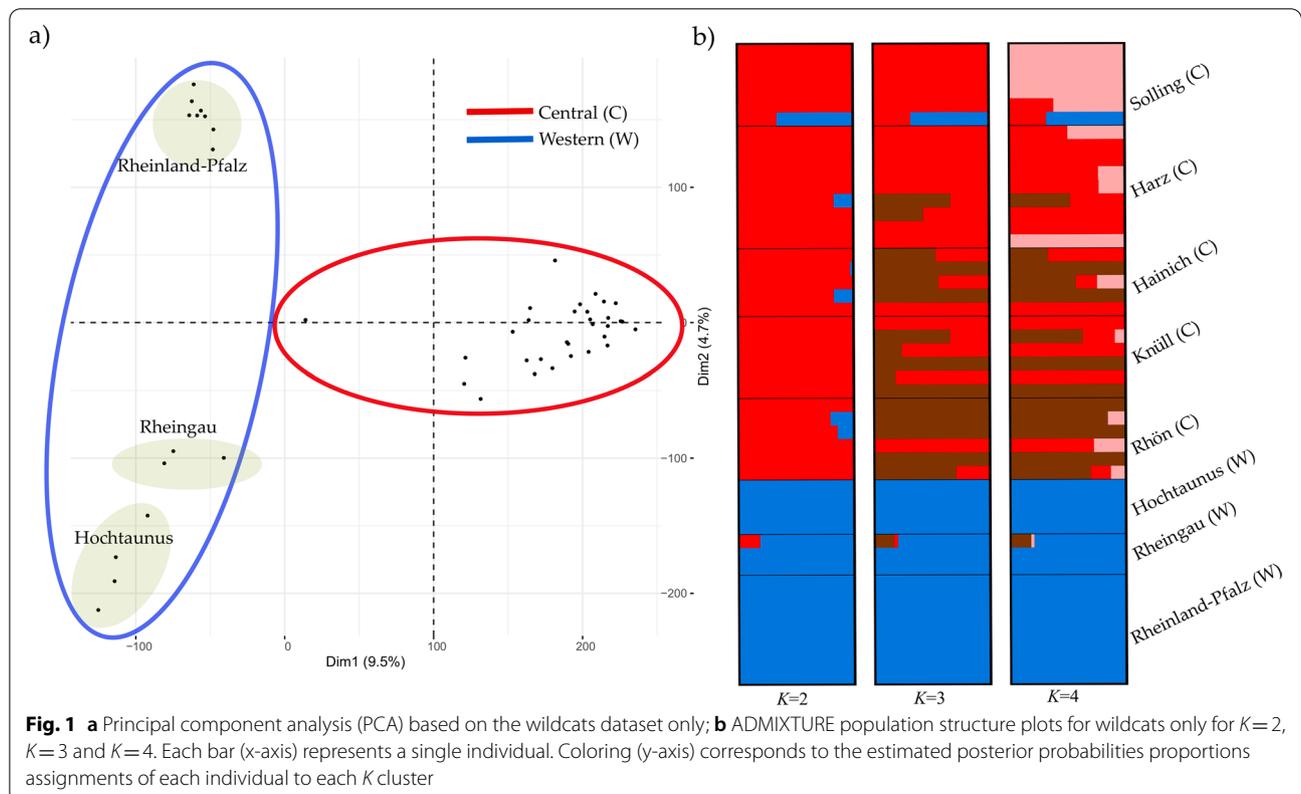
We generated whole-genome sequencing data for 47 wildcat and 16 domestic cat individuals of a mean coverage of 18-25X. A total of 38,090,943 SNPs (single nucleotide polymorphisms) were found in the full data set containing wildcat and domestic cat, and 28,618,073 SNPs after passing filtering steps.

ADMIXTURE showed a clear structure for the dataset including domestic cat and wildcat ($K=2$, Figure S1, Supporting Information) with a log-likelihood value = -71,316,531.6 and lowest cross validation (CV)-error = 0.312. When only wildcat were analyzed, both ADMIXTURE and principal component analysis (PCA) methods found statistical support for two clusters ($K=2$, Fig. 1a) with a log-likelihood value = -17,823,710.6 and lowest CV-error = 0.432. The PCA approach clustered individuals according to the a priori location assignment (Fig. 1b). PC1 explained 9.5% of the total variance and split Western and Central wildcats, whereas PC2, explaining 4.7% of the total variance, showed a split between Rheinland-Pfalz individuals and, the Rheingau and Hochtaunus region, which are separated by

the Rhine river. Our results do not show an apparent clustering within the Western cluster group. Individual “FB021A” from Solling appeared as an intermediate individual and did cluster neither with Western nor Central wildcats.

Historical demography and scenarios of divergence

Both analyses, PSMC and MSMC2, recovered clearly different demographic histories for the two species and much lower effective population size (N_e) for wildcats in comparison to domestic cats. The PSMC analysis showed no significantly different demographic trajectories between Central and Western wildcat individuals, indicating the existence of a single demographic population for most of the time. Results showed a steep N_e growth of wild cats in the Middle-Pleistocene (ca. 400 thousand years ago (kya)) reaching the peak in the Upper-Pleistocene with an average N_e of 70,000 individuals (ca. 110 kya), and followed by a continuous decrease since then (Figure S2a and, Figure S3 for bootstrap ranges). The MSMC2 analysis yielded extremely high N_e values which we do not take as face values but rather as a relative measure. However, there was indication of a steep and sudden drop after the postglacial demographic increase a few hundred years ago. Results also showed the different postglacial trajectories of the two wildcat populations



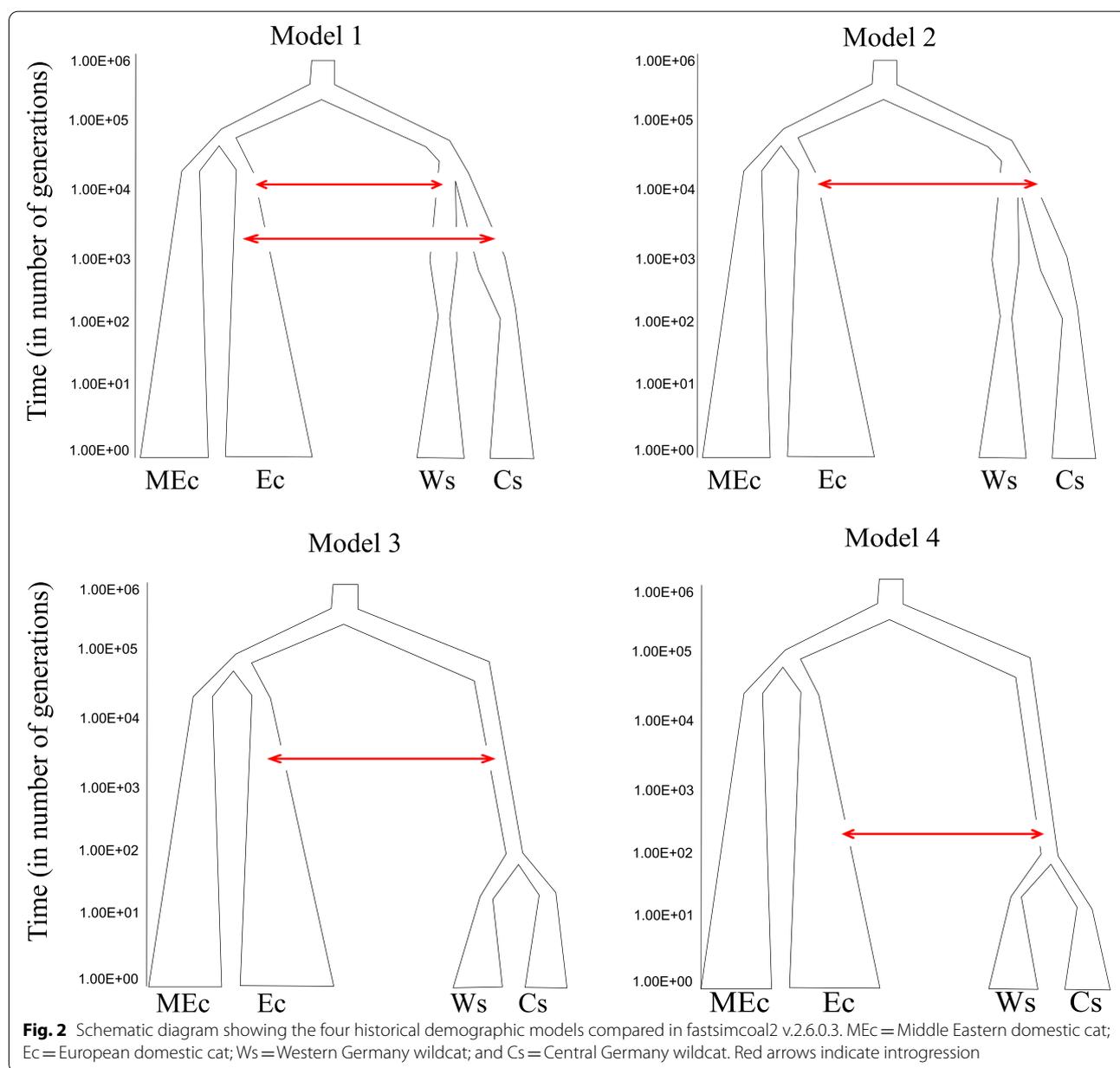
and the relative difference between timing and drop in N_e (Figure S2b).

The Approximate Bayesian Computation (ABC) demographic analysis recovered a proportion of accepted simulations for all demographic models tested. Demographic model 3, which hypothesized a late split of western and central wildcat populations and early introgression, received the lowest support of all. The best-supported demographic model was model 4 (Fig. 2), with 95% of accepted simulations of (Table 1).

Model 4 suggested a late introgression event (ca. 600 generations ago) and a recent split (ca. 100 generations

Table 1 Model selection from fastsimcoal2 analyses. Bold indicates the demographic scenario that received the highest support. *pp*'s = posterior probabilities

| Demographic model | % of simulations accepted (<i>pp</i> 's) |
|-------------------|---|
| M1 | 0.0325 |
| M2 | 0.0075 |
| M3 | 0.005 |
| M4 | 0.955 |



ago) of western and central wildcat populations (Table S1). Second best model (Model 1, 3% of accepted simulations, Table 1) suggested also an early split (ca. 3500 generations ago) of western and central wildcats and differential introgression. Together, models with a very recent split of western and central wildcat populations received >95% support. The best tested model (Model 4) provides a good fit to the data (Figure S4, Supporting Information).

Inbreeding: runs of homozygosity

The proportion of the genome that was contained in runs of homozygosity (F_{ROH}) ranged between 0.004 and 0.2 (Figure S5 in Supp. Information). There were no significant differences in ROH measurements between the

West and Central populations, nor significant regional structuring ($P=0.39, 0.72$). Analysis of long ROH (> 2 Mb and >5 Mb), which can detect recent rather than distant inbreeding, identified 1,771 ROH segments in 45 individuals and 57 segments in 30 individuals, respectively. Thus, while at least some degree of recent inbreeding was detected in 63% of individuals, it was not tied to population structuring within and between populations.

Selection analyses and functional enrichment analyses

Empirical and simulated F_{ST} distributions between Western and Central wildcats were similar, with simulated distribution having a higher kernel density of occurrences (Fig. 3a). We found 64 genes with F_{ST} values above the 99% quantile of the simulated F_{ST} . These genes were

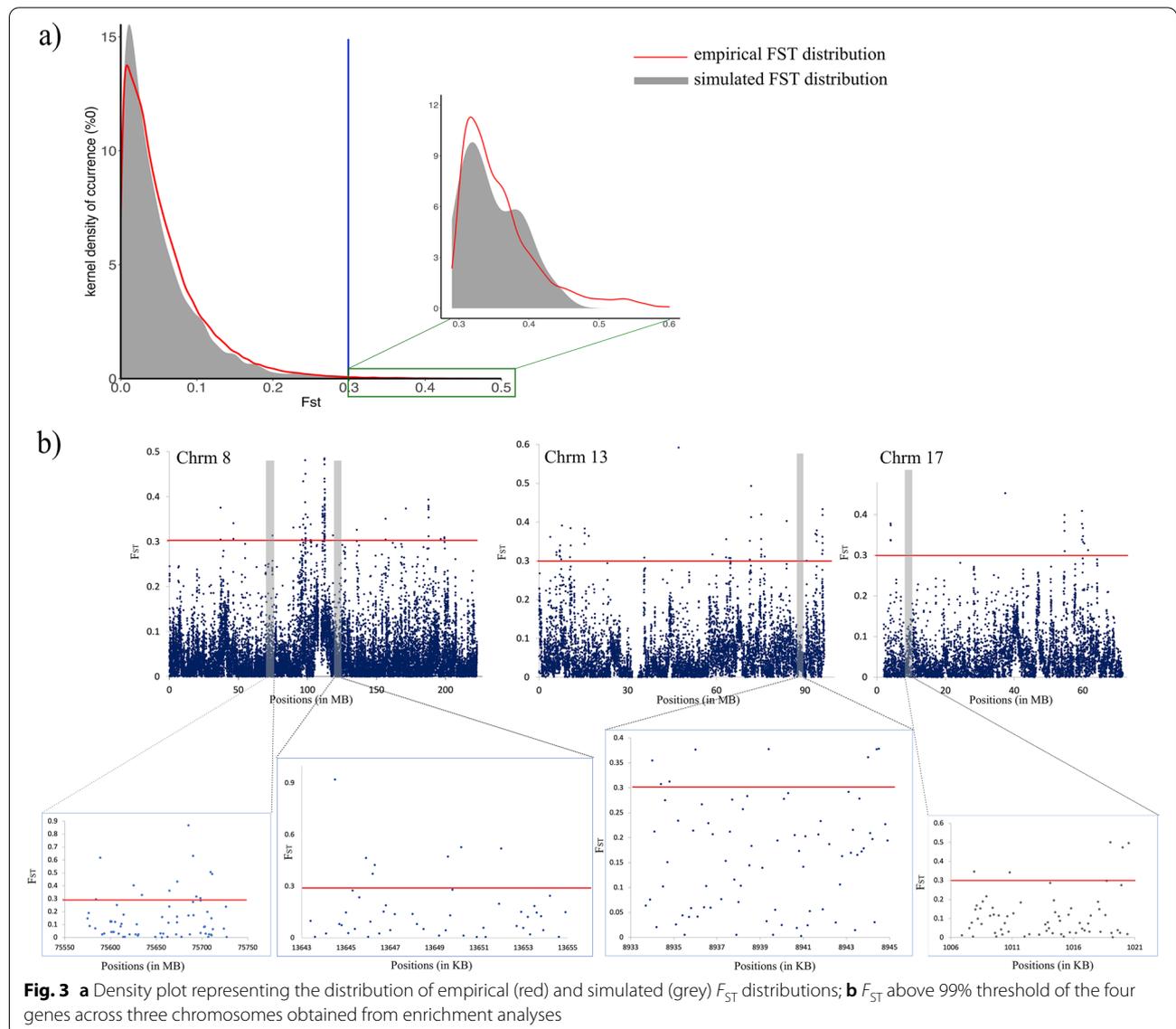


Fig. 3 a Density plot representing the distribution of empirical (red) and simulated (grey) F_{ST} distributions; b F_{ST} above 99% threshold of the four genes across three chromosomes obtained from enrichment analyses

therefore considered as potential candidate genes under selection. Associated with the 64 candidate genes, 16 gene ontology (GO) terms were significantly enriched for biological processes (BP -Table S2, Supporting Information). After applying the FDR correction only two GO terms remained statistically significant.

Four genes associated with two significantly overrepresented top GO terms (Fig. 3b and Table S2) were associated with kynurenine metabolic and tryptophan catabolic processes. In particular, these genes were involved in the coding of cysteine-S-conjugate beta-lyase 2, kynureninase, kynurenine aminotransferase 1, and kynurenine 3-monooxygenase (tryptophan-kynurenine-serotonin pathway in Figure S6, Supporting Information).

Haplotype relationships

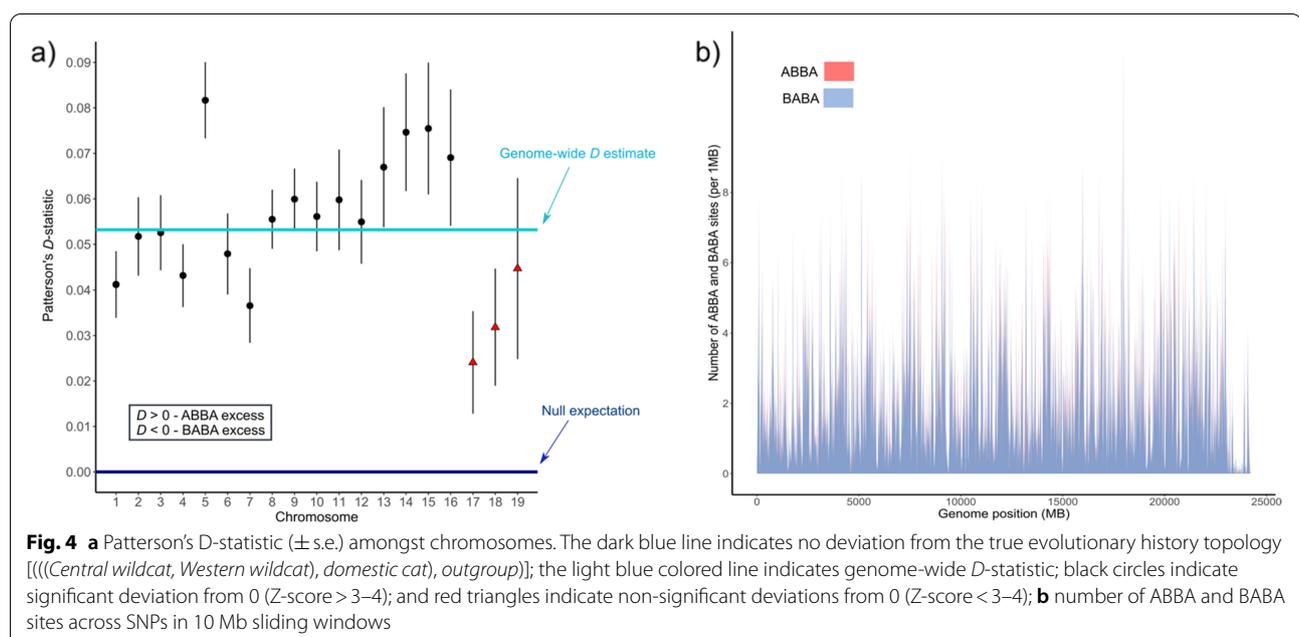
Sequence length (in bp) and number of polymorphic sites for the four genes identified in the selection analyses were 77,142 and 34, 111,195 and 59, 35473 and 58, and 57,619 and 33, respectively. A total of 17, 18, 0, and 21 unique haplotypes were found for genes 1–4 respectively. The haplotype network diagrams for genes 1–3 show a clear grouping of wildcat vs domestic cats haplotypes, except for gene four, for which there is not that apparent structure. Almost no haplotypes were shared between wild and domestic cats. Overall, the number of haplotypes shared amongst the three groups was insignificant, while shared haplotypes between Central and Western wildcats were more frequent (Figure S7 in Supp. Information).

Introgression between FS and FC

The genome-wide ABBA-BABA test identified a very slight excess of the ABBA pattern (Fig. 4a - $D=0.0532$), indicating introgression between Western wildcat and domestic cat. The Z score ($Z=24.84$) and p -value ($p\text{-value}=3.302708\text{e-}136$) were significant for the D estimation. The proportion of the genome introgressed between Western wildcat and domestic cat was relatively low ($f=0.08855515$; 95% confidence interval 0.0812–0.0959), however, the proportion of the genome introgressed between Central wildcat and domestic cat remains unknown. ABBA and BABA site counts across the genome were similar (Fig. 4b) and the chromosome level analysis also identified an excess of ABBA pattern (Fig. 4a - D estimation for all chromosomes >0). However, Z score was not significant ($Z<3-4$) for chromosomes 17–19 (Table S3 in Supp. Information).

Discussion

In this study, we used WGS data for the first time to investigate the demographic history of German wildcats and the potential role of hybridization with domestic cat or other anthropogenic effects on population differentiation. Our analyses showed that the split of western and central groups of FS in Germany was quite recent, probably due to anthropogenic impact, rather than ancient historical events as previously believed and most likely not due to differential introgression.



Phylogeographic structure and historical demography of German wildcats

Our results from the population structure analyses clearly showed two groups of German wildcats that corresponded to western and central locations, which are in agreement with previous studies that have used limited genetic information from microsatellites or reduced SNP panels [18, 21–23, 33]. The PCA clustering analysis further separated the western group into three subgroups, which corresponded to Rheinland-Pfalz (west of Rhine River), Rheingau and Hochtaunus locations (east of Rhine). In contrast to [18, 24], our analyses did not recover the subgroups Rheingau and Hochtaunus, as expected due to major traffic infrastructures which act as gene flow barriers. In accordance with [18], our PCA analysis showed no substantial substructure in the central cluster (Fig. 1a and Figure S1).

The model-based clustering method showed an unexpected highly homogeneous western group, with just one genetic pool recovered even at higher K s (Fig. 1b and Figure S1). One potential explanation might be the high forest cover and low grades of habitat fragmentation in the western part of the species' range in Germany that still led to substantial genetic exchange. However, we cannot discard the possibility that ADMIXTURE did not reveal substructure in the western cluster due to the small sample size per location. In contrast, central wildcats showed considerable differentiation when only wildcats are analyzed (Fig. 1b), which likely corresponds to geographic differentiation during the strong bottleneck period in the late 19th and early twentieth century [20]. In addition, founder effects resulting from recent expansions and reintroductions, for example in the Rhön Mountains [19], might have contributed to the currently observed spatial genetic structuring. Our study shows that wildcats in Germany do not show signs of substantial inbreeding or genetic depletion, which is likely due to the ongoing recovery after cessation of the massive persecution since the second half of the twentieth century [20]. In contrast, relatively long ROHs values were found for single individuals (Figure S5b), which suggest that inbreeding may occur occasionally in all populations.

Effective population sizes (N_e) for western and central German wildcat have experienced subtle changes, but generally, western and central wildcats present similar N_e . There is also no population separation recovered by the PSMC analyses for the period where this analysis yielded sufficient resolution i.e. before 10000 years ago (Figure S2a). As seen for other organisms, the N_e peak for wildcats occurred towards the end of the last interglacial period (ca. 110 kya), which was a period of the Late Pleistocene characterized by wetter and warmer environmental conditions [34–36]. This period also coincides

with the upper limit of the Middle to Late Pleistocene where an increase of small-sized carnivores is observed as a consequence of the climate oscillations [37]. The N_e inference using MSMC2 is most likely biased by phasing errors and results should be taken cautiously. It has been shown that phasing errors in haplotype-based inference methods, such as MSMC, break up identity-by-state tracts in closely related haplotypes which leads to inferring large recent effective population sizes [38].

The Approximate Bayesian computation analysis supported a relatively recent introgression between wildcats and European domestic cats, and a only very recent split of western and central wildcat populations around 100 generations ago (Model 4, Fig. 2). The late introgression suggested by the best fit model ca. 700–1,000 generations ago occurred not directly upon the time of the domestic cat introduction by the Romans, but rather during the medieval spread of cats as primary domestic mousers. Our results indicated that the divergence of western and central wildcat groups occurred rather recently. In principle, a population split can be due to loss of habitat (including land-use changes) or a demographic decline, leading to extirpation of local populations and eventually isolation of relic populations in refugia [39]. Even though overall wood cover in Europe did not change significantly since 1600, changed forestry practices might have negatively impacted wildcats demography [40]. In particular the shift from unmanaged or coppiced deciduous woods to managed coniferous high stands from the mid-nineteenth century onwards [40] probably meant a significant habitat reduction for the species [41–43]. The most likely reason for the observed split, however, is a demographic decline due to anthropogenic hunting pressure. Persecution of wild cats probably increased with the widespread introduction of effective hunting shotguns in the second half of the eighteenth century [44], augmented by societal changes in hunting privileges [45]. Both the broad availability of an effective hunting weapon for this small elusive carnivore as well as the increase in the number of hunters by permitting bourgeoisie and peasant hunting likely led to a rapid decline of the species, perceived as pest and competitor for prey.

This finding of a very recent, anthropogenically induced split between western and central wildcats is in line with results based on museum material of wildcats from Germany. A recent study [20] showed that historic material collected before World War II from across Germany did not display any signals of population differentiation. This is in line with our findings and strongly suggests that the emergence of the central wildcat population is a consequence of historic persecution and a resulting severe bottleneck and subsequent re-expansion in recent decades [20]. Interestingly, previous studies based only on

microsatellite markers had erroneously concluded that the divergence between western and central wildcats derive rather from more ancient, natural processes of glacial isolation, and suggested to treat both lineages as divergent clades [21, 25]. Correcting this assumption has important conservation implications for this endangered species (see below).

Functional basis of population divergence

Some of the highly differentiated genes are involved in the tryptophan-kynurenine-serotonin pathway, in particular in the catabolism of secondary metabolites and production of nicotinamide metabolism (NAD). The fact that the differentiated genes are ultimately involved in the serotonin production might have important behavioral implications on the wildcats. Tryptophan is the metabolic precursor to serotonin, which is a neuromodulator involved in the regulation of several behavioral processes, including aggression, mood, and stress susceptibility and -resistance in both vertebrates and invertebrates [46, 47]. In fact, supplementing diet with tryptophan is widely used in domestic and livestock animals to modulate aggressiveness, hyperactivity and stress recovery [48–51].

It is widely accepted that human activities may negatively impact and have fitness consequences to wildlife, including stress, disease and reproductive success [52–54]. With the evidence from our study, we were not able to identify whether the production of serotonin is actually upregulated or downregulated in the wild cats. Neither could we determine with the data at hand whether one of the two German wildcat clusters might display a higher stress tolerance. Therefore, investigating different wildcat genotypes regarding stress tolerance/resistance level would shed more light into this issue. We offer two potential, mutually non-exclusive hypotheses to the observed differentiation at these genes in German wildcat populations: 1) reintroductions in one of the two populations have substantially supplemented populations with animals that have been previously raised and kept in captivity. Long-term close human contact in zoos might have led to the selection of stress resistance in these wildcats – more heritably stress resistant individuals might have better survived and reproduced in captivity, 2) an increased frequency of human impact within wildcat habitats (e.g. leisure activities like hiking, mountain biking, geochaing etc.) during the past decades might have posed selection pressures for wildlife to cope with stress derived from such interactions. Therefore, higher tolerance to stress can pose a significant adaptive advantage over animals that do not cope well with the stress caused by increased human contact. This hypothesis is in line with strong evidence of wildcat recovery and expansions across the structured anthropogenic habitats and

mosaic-like landscapes [18, 19, 55] and other medium- and large-sized carnivores in Central Europe [56, 57].

Introgression between wild and domestic cats, and its role on population divergence

We found some evidence of genome-wide introgression between German wildcats and domestic cats. However, the proportion of genome introgressed is relatively low, which supported previous evidence of low levels of introgression in German wildcats [21–23]. The relatively low levels of introgression observed in German wildcat populations might be due to the persistence of large forest patches in low mountainous regions [22], which may minimize both the opportunity and the necessity of interactions between wild and domestic congeners. In addition, we did not observe differential introgression for different wildcat populations. Deviations from the true bifurcating evolutionary history of the taxa are rather similar (Fig. 4b), indicating that gene flow between domestic cats and either Western or Central wildcats was and is equally likely, with no prevalence towards any of the groups. The chromosome level analysis, however, suggested a slightly increased level of introgression between Western wildcats and domestic cats (excess of ABBA sites ($D > 1$); Fig. 4a), which may be explained by the earlier contact of Western wildcats with domestic cats as the introduction of domestic cats likely happened first in south West Germany, which was part of the Roman Empire [27].

The relationship amongst haplotypes for each of the four highly differentiated genes reveals a distinct structure that corresponds to the different geographical groups defined in the analyses, a pattern which is neatly congruent with the population structure analyses (Figure S1, Supp. Information). Genetic drift, mutation and selection, but also by introgression from between congeners [58], shape the evolution of genomes in natural populations. However, in this case introgression from *FC* does not seem to play an important role in the differentiation of these genes between Western and Central wildcat populations, as shown by the very few haplotypes shared between domestic cat and the two wildcat populations. This might be due to the fact that introgression started relatively early in German populations and is generally very low [18, 21, 22], and overall, we do not see significant admixture in the four highly differentiated genes as it would be expected if introgression had contributed to population differentiation.

Conservation implications

The observed ongoing spread of the European wildcat in Germany [18, 20] has been accompanied by large scale species conservation projects, aiming at facilitating the

spread of the species by reducing habitat fragmentation within the patchy, mosaic like distribution of suitable habitats within the anthropogenically modified cultural landscape. For this, forest corridors were planted in different regions of the wildcat's distribution to facilitate wildcats and other species disperse between isolated forest patches and migrate between different densely forested low mountain regions [59]. While these conservation attempts have gained considerable public attention, several related questions remain unanswered to date, such as potential negative effects of connectivity measures, e.g., related to potentially merging naturally isolated phylogenetic lineages or if higher landscape connectivity may facilitate the spread of introgression from domestic cats.

While our findings may question the need for landscape migration corridors to facilitate gene flow between isolated habitats, they may still be valuable on a local perspective given the significant effects of and high abundance of effective landscape barriers, such as roads or settlements in the landscapes of Germany [60]. Given the recent anthropogenic divergence between west and central wildcat populations, there is currently no evidence against reconnecting them. Also, the relatively low proportion of introgression observed from domestic cats in the wildcat genomes indicated that an undesired spread of domestic cat genes into the wildcat population due to facilitated gene flow appears rather unlikely at the moment.

Given the fact that so far only a few corridors have been planned and at least partially implemented over rather short spatial distances [59], it appears unlikely that the observed high connectivity and genetic diversity as well as relatively low level of inbreeding detected in the wildcat genomes are consequences of those conservation actions. The observed patterns rather seem to be the consequence of the ongoing population increase and expansion after protection of the wildcat in the 1940s, as it has been documented in several studies [18–20, 61].

Conclusions

Within Europe's cultural landscapes, humans have had a major influence on wildlife for centuries, but the underlying genomic consequences of human-mediated activities on wildlife are still poorly understood. Although the use of population genomics applied to wildlife conservation is relatively new, it has been shown to have a great potential for the understanding of species ecology and biology as well as to inform species conservation management [11]. Our data suggest an evident impact of anthropogenic pressures on population differentiation and adaptive selection on German wildcat populations in the last 200 years. Future research should couple genomic

evidence, such as the ones presented in this study, with behavioural evidence (e.g. identification of wildcat phenotypes that are more stress resistant) in order to test the hypotheses proposed above, and gain more insights on how short term anthropogenic activities and conservation strategies impact wildlife adaptive selection processes.

Methods

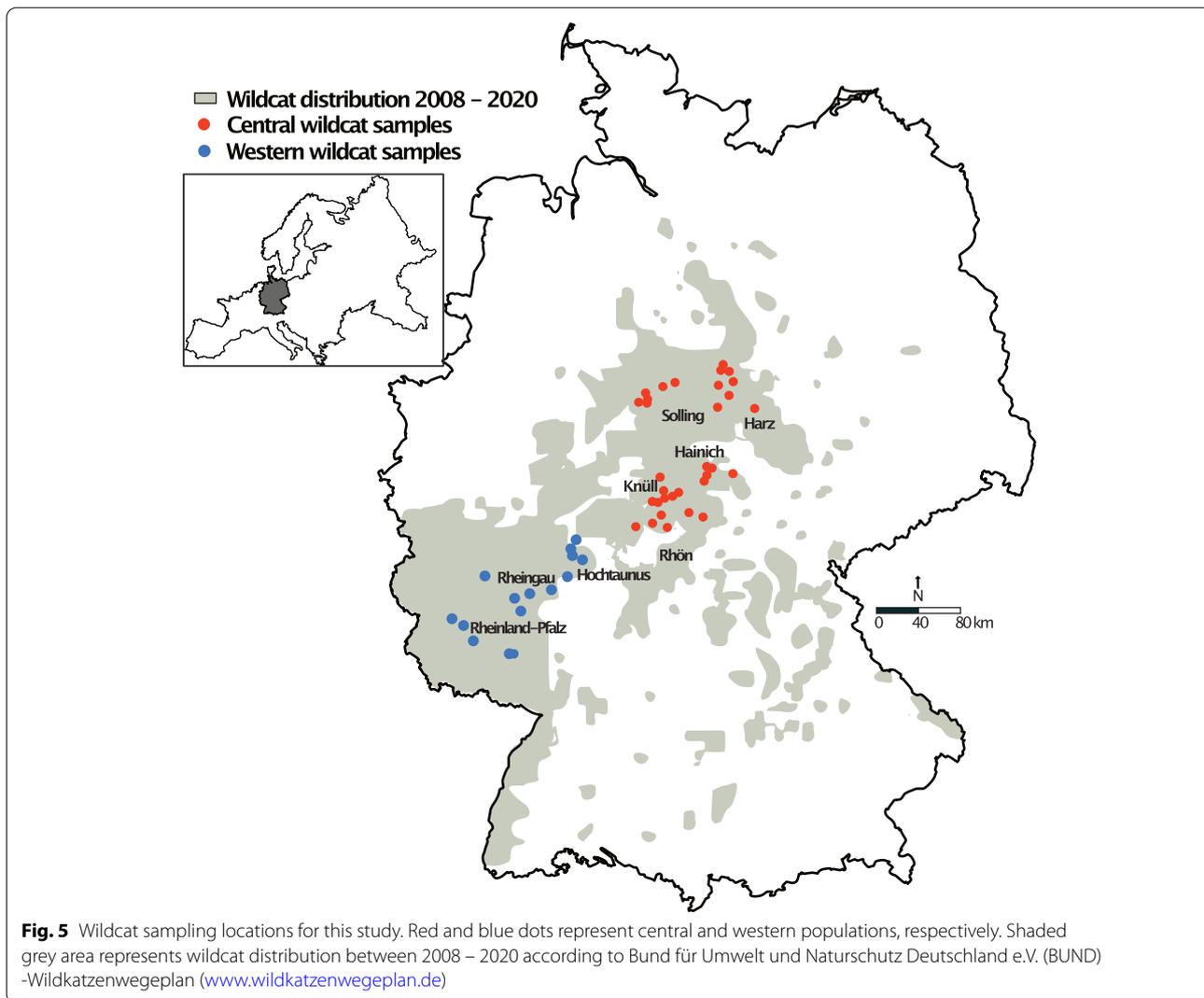
Sampling, sequencing, filtering and variant calling

We whole-genome sequenced 47 *FS* individuals from across the species' distribution in Germany (Fig. 5) and, 16 *FC* individuals from around the world. In addition, we obtained fastq files for 21 domestic cat individuals from the ENA project PRJNA343389. Our final taxon sampling included 84 *Felis* individuals (Table S4 in Supp. Information). BGI Genomics conducted DNA isolation and library preparation. Sequencing was performed on a BGISEQ-500 platform, which generated 100 bp paired end reads per individual, and a mean coverage of 18–25X. Quality of raw whole genome sequence data of all 84 samples was checked using FastQC v0.11.5 [62] in combination with MultiQC v1.5 [63]. Quality checks indicated that reads were adapter free and were then mapped to the *Felis catus* v9.0 reference genome (European Nucleotide Archive (ENA) accession number: GCA_000181335) using BWA mem v0.7.15 [64] with default settings, and the $-M$ option for Picard (Broad Institute 2019) compatibility (for downstream duplication marking. SNPs were called with Platypus v1.0 [65]. Bases with quality scores below 30 and reads with mapping quality below 30 were ignored, and only variants with at least 6 reads were kept.

We filtered for biallelic variants passing all filters and pruned all SNPs for Linkage disequilibrium (LD) with a squared correlation coefficient of more than 0.5 using bcftools v1.9 [66].

Population structure analyses

We used ADMIXTURE version 1.3.0 [67] to infer ancestry proportions based on maximum likelihood estimations using an increasing number (K), from $K=1$ to $K=15$, with default settings and a tenfold cross validation procedure. We performed a second ADMIXTURE analysis for a subset that included only German wildcat individuals, $K=1$ to $K=10$ using the same settings as the first analysis. We subsequently performed a 10 runs analysis for $K=2$, the K that received the highest support for German wildcat only. Results from both analyses were plotted with PONG [68]. In addition, we conducted a principal component analysis (PCA) on 1,899,534 unlinked SNPs on the two datasets, all *Felis* individuals and just German wildcat individuals, using the R package *Factoextra* v.1.0.7 [69].



Demographic history inference

We implemented a pairwise sequentially Markovian coalescent (PSMC, [70]) model in order to reconstruct the demographic history of wildcat in Germany. We created fastq sequences for eight unadmixed individuals selected based on ADMIXTURE results, including 2 of each of the following groups: western wildcats, central wildcats, western domestic cats and eastern domestic cats. We used SAMTOOLS v1.9 “mpileup” command [71] to filter for bases with a base quality of 30 and minimum coverage of SNPs of 10 as recommended by [72]. PSMC was run with 25 iterations and the upper limit of TMRCA was set to 5, initial ρ/θ value to 5 and N_e was inferred across 64 interval times ($2 \times 3 + 25 \times 2 + 4 + 4$) and calculated bootstrap ranges for each individual. Results were scaled with a mutation rate of $\mu = 1 \times 10^{-8}$ per base pair and generation, assuming a generation time of two years [33]. We also implemented a multiple sequentially

Markovian coalescence approach (MSMC2, [73] using the same set of individuals and groups as for the PSMC analysis). MSMC2 is a more advanced method which increases accuracy of the N_e history (especially inference of the recent past) and uses phased haplotype data of single chromosomes; therefore we separated each of the autosomal chromosomes and used Beagle v.5.4 [74] for phasing the genotypes using default parameters and window size of 10 kb. We used the generate_multihetsep.py (MSMC-TOOLS package; <https://github.com/stschiff/msmc-tools>) to generate the input files for MSMC2 for each chromosome. We used the same groups of unadmixed individuals, generation time, mutation rate and interval times as for the PSMC analysis. For the MSMC2 runs we used the option to skip sites with ambiguous phasing (-s) as recommended by instructions.

We performed a demographic modelling analysis where we defined four groups. Based on the clear

population structuring from the ADMIXTURE analyses results we included central German *FS* and western German *FS*. The Middle East *FC* individuals were included in the analysis to polarize derived vs ancestral alleles, and the European *FC* as potential source of introgression with the *FS*. We designed four competing demographic models taking into account PSMC results and considering three criteria: 1) temporally different split time between central and western *FS*; 2) different temporally gene flow between European *FC*, central *FS* and western *FS*; and 3) different temporally introgression between European *FC* and *FS*. Based on these criteria we designed four models as follows: Model 1) depicts an early split of Western and Central German wildcat populations and differential introgression in wildcat; Model 2) depicts an early split of Western and Central German wildcat populations and late introgression; Model 3) depicts a late split of Western and Central German wildcat populations and early introgression; and Model 4) depicts a late split of Western and Central German wildcat populations and late introgression (Fig. 2). Here, we defined introgression as a one way admixture historical event (from domestic to wildcats). Detailed description of prior parameters for each model is given in the Supplementary Methods. We selected eight individuals per group based on GC% and coverage, except for near east *FC* for which just five samples were available. We conducted coalescent simulations of the multidimensional *site-frequency spectrum* (SFS) as summary statistics. In order to minimize the effects of selection on demographic inference, we included only neutrally and no linked evolving SNPs at fourfold degenerated sites in the genome by using tbg-tools v.0.2 (<https://github.com/Croxa/tbg-tools>) for selection. The selected SNPs were pruned for linkage disequilibrium (LD) with PLINK v 1.9 [75, 76], applying an r^2 threshold of 0.1. We obtained the observed SFS of the unlinked, neutral SNPs using easySFS (<https://github.com/isaacovercast/easySFS>). We ran 10,000 simulations with fastsimcoal2 v.2.6.0.3 [77] for each model based on the observed multidimensional SFS (input files for the fastsimcoal2 analyses given in the Appendix). Simulated and observed SFS for each model were then compared using an Approximate Bayesian Computation (ABC) approach with the R package *abc* [78].

ROH analyses

Runs of homozygosity (ROH) were identified using PLINK v 1.9 [75, 76]. We used the `--geno` option in PLINK before running the ROH analyses to keep only SNPs with a 90% genotyping rate (10% missing). There is a lack of consensus in the literature to whether SNP data should be pruned by LD and/or minor allele frequency (MAF). We followed Meyermans et al. [79]

recommendations and skipped MAF pruning but tested different homozygosity window SNP and thresholds. In total we ran 3 analyses using the following options: 1) `-homozyg -homozyg-window-snp 20 -homozyg-window-het 1 -homozyg-window-missing 5 -homozyg-window-threshold 0.05 -homozyg-snp 20 -homozyg-het 1 -homozyg-density 100 -homozyg-kb 1000 -homozyg-gap 1000`; 2) we increased the `-homozyg-window-snp` to 50; and 3) `-homozyg-window-snp 20` and `-homozyg-window-threshold 0.25`. We estimated the fraction of the genome in ROH (F_{ROH}) as the total length of ROH per individual divided by the length of the genome. We evaluated significance level of the estimate using a pairwise t-test with Bonferroni correction.

Fixation index analyses F_{ST}

We estimated the genome-wide fixation index (F_{ST} ; [80]) for wildcat only (Western vs Central) using a non-overlapping 10-kb window with VCFtools 0.1.17 [81]. In order to get an indication of significant genes (outliers) under selection, we used parameters of the best-fit inferred demographic scenario to simulate 10-kb sequences using fastsimcoal2 v.2.6.0.3 [77] to infer F_{ST} distribution, under the assumption of no selection. We used a custom python script to extract genes for which any of the outliers 10-kb windows above the 99% quantile of the simulated F_{ST} distribution fall within the genomic coordinates of genes of the annotated genome of *FC*. Resulting genes were the candidates for functional enrichment analyses.

Functional enrichment analyses

We ran a gene ontology (GO) term enrichment analysis on candidate genes that fell above the 99% F_{ST} threshold of simulated F_{ST} . GO terms were annotated to all protein sequences using InterProScan v 5.39–77.0 [82], resulting in 45,297 GO terms which served as the reference for the functional enrichment analyses. We used 15,224 GO-annotated genes as universe. Analysis was carried out using the R package topGO v.2.42 [83] for the category biological processes (BP). We used the "weight01" algorithm and Fisher test for statistical support. We retained GO terms with a p -value ≤ 0.05 and used the Benjamini–Hochberg method [84] for p -value correction with a FDR (false discovery rate) level of 5%.

Haplotype networks

In order to visualize and interpret individual relationships at the population level, we built haplotype networks, which can be more appropriate than hierarchical tree format for intraspecific analyses [85, 86]. We selected the four genes that corresponded to the 2 top GO terms from the enrichment analysis and extracted the DNA consensus haplotypes sequences from the *Felis catus* v9.0

reference genome for all cats included in our sampling using VCFtools 0.1.17 [81] and bcftools 2.4.37 [71]. Haplotype DNA sequences for all 84 individuals were phased with DnaSP v6 [87] and aligned using MAFFT v7 [88, 89]. We built haplotype networks for the four regions independently using the TCS algorithm [90] implemented in PopART 1.7 [91]. We coded individuals in 3 groups, corresponding to domestic cats, Central German wildcat and Western German wildcat.

Introgression test

We tested for genome-wide introgression between wildcat and domestic cat with the ABBA-BABA test [92] using Patterson's *D*-statistic [93, 94]. We defined the groups as follows: P1 corresponds to Central wildcat, P2 corresponds to Western wildcat, P3 corresponds to European domestic cat. We used five Middle Eastern domestic cats as outgroup, and used the following species tree topology as null expectation: BBAA [(((P1,P2),P3),O)]. We tested 2 hypotheses to look at differential introgression as follows (Figure S8 in Supp. Information): 1) ABBA pattern would indicate gene flow between Western wildcat and domestic cat [(((P2,P3),P1),O)], and 2) BABA pattern gene flow between Central wildcat and domestic cat [(((P1,P3),P2),O)]. We calculated the genome-wide and chromosome level *D*-statistic using the genomics general toolkit (https://github.com/simonhmartin/genomics_general; [92]) and the proportion of the genome introgressed (*f*). To test for the consistency of the estimation, we computed the standard deviation of *D*-statistic with the block jackknife procedure with 95% of the data, and the standard error, *Z* score and *p*-value for the test. In addition to this, we calculated *D*-statistic in 100-kb non-overlapping windows.

Abbreviations

ABC: Approximate Bayesian Computation; BP: Biological processes; CV-error: Cross validation error; FC: *Felis catus*; F_{ROH} : Proportion of the genome contained in runs of homozygosity; FS: *Felis silvestris*; F_{ST} : Fixation index; GO terms: Gene ontology terms; kya: Thousand years ago; MAF: Minor allele frequency; NAD: Nicotinamide; N_e : Effective population size; PCA: Principal component analysis; PSMC: Pairwise sequentially Markovian coalescent; ROH: Runs of homozygosity; SFS: *site-frequency spectrum*; SNP: Single nucleotide polymorphism; TMRCA: Time to most recent common ancestor; WGS: Whole-genome sequencing.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08930-w>.

Additional file 1. Supplementary Results, Supplementary Methods and Appendix A.

Acknowledgements

We thank the BUND Germany and several other long-term collaborators for sample provision and long-term support, and L. Chueca and T. Schell for bioinformatics support.

Authors' contributions

MP and CN designed the study. DS performed data preparation and SNP discovery. MENB performed rest of data analyses. SM assisted with ROH analyses interpretation. KK and SS contributed with Figures. MENB wrote the manuscript and all authors contributed to the preparation of the final version and approved it.

Funding

Open Access funding enabled and organized by Projekt DEAL. This was funded by funded by LOEWE-TBG initiative.

Availability of data and materials

Newly generated whole genome individual sequencing data is available at European Nucleotide Archive (ENA) project number: PRJEB40421 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB40421>). fastq files for 21 domestic cat individuals can be found under ENA project PRJNA343389 (<https://www.ebi.ac.uk/ena/browser/view/PRJNA343389>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Molecular Ecology Group, Senckenberg Biodiversity and Climate Research Centre (SBIK-F), 60325 Frankfurt am Main, Germany. ²Division of Evolutionary Biology, Faculty of Biology, Ludwig Maximilian University of Munich, Planegg-Martinsried 82152, Munich, Germany. ³Centre for Wildlife Genetics, Senckenberg Research Institute and Natural History Museum Frankfurt, 63571 Gelnhausen, Germany. ⁴European Wildcat Monitoring, Bund Für Umwelt Und Naturschutz, Rheinland-Pfalz, 55118 Mainz, Germany. ⁵LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), 60325 Frankfurt am Main, Germany. ⁶Institute for Molecular and Organismic Evolution, Johannes Gutenberg University, 55128 Mainz, Germany.

Received: 20 May 2022 Accepted: 7 October 2022

Published: 18 October 2022

References

1. Bull JW, Maron M. How humans drive speciation as well as extinction. *Proc R Soc B Biol Sci.* 2016;283:20160600.
2. Sala O, Chapin III FS, Armesto JJ, Berlow E, Bloomfield J, RH D, et al. Biodiversity - Global biodiversity scenarios for the year 2100. *Science.* 2000;287:1770–4.
3. WWF. Living Planet Report 2016, Risk and resilience in a new era. WWF International. Gland, Switzerland; 2016.
4. Palumbi SR. Humans as the world's greatest evolutionary force. *Science.* 2001;293:1786–90.
5. Hendry AP, Gotanda KM, Svensson EI. Human influences on evolution, and the ecological and societal consequences. *Philos Trans R Soc B Biol Sci.* 2017;372:20160028.
6. Pelletier F, Coltman DW. Will human influences on evolutionary dynamics in the wild pervade the Anthropocene? *BMC Biol.* 2018;16:7.
7. Allendorf FW, Hard JJ. Human-induced evolution caused by unnatural selection through harvest of wild animals. *Proc Natl Acad Sci.* 2009;106(Supplement 1):9987–94.

8. Fugère V, Hendry AP. Human influences on the strength of phenotypic selection. *Proc Natl Acad Sci*. 2018;115:10070–5.
9. Hendry A, Farrugia TJ, Kinnison MT. Human influences on rates of phenotypic change in wild animal populations. *Mol Ecol*. 2008;17:20–9.
10. Alberti M, Correa C, Marzluff JM, Hendry AP, Palkovacs EP, Gotanda KM, et al. Global urban signatures of phenotypic change in animal and plant populations. *Proc Natl Acad Sci*. 2017;114:8951–6.
11. Hohenlohe PA, Funk WC, Rajara OP. Population genomics for wildlife conservation and management. *Mol Ecol*. 2021;30:62–82.
12. Allendorf FW, Leary RF, Spruell P, Wenburg JK. The problems with hybrids: setting conservation guidelines. *Trends Ecol Evol*. 2001;16:613–22.
13. Rhymer JM, Simberloff D. Extinction by hybridization and introgression. *Annu Rev Ecol Syst*. 1996;27:83–109.
14. McFarlane R, Sleight A, McMichael T. Synanthropy of wild mammals as a determinant of emerging infectious diseases in the Asian-Australasian Region. *EcoHealth*. 2012;9:24–35.
15. Streicher JP, Ramesh T, Downs CT. An African urban mesocarnivore: navigating the urban matrix of Durban. *South Africa Glob Ecol Conserv*. 2021;26:e01482.
16. Stahl P, Artois M. Status and conservation of the wildcat (*Felis silvestris*) in Europe and around the Mediterranean rim. Council of Europe. Strasbourg; 1995.
17. Yamaguchi N, Kitchener A, Driscoll C, Nussberger B. *Felis silvestris*. The IUCN Red List of Threatened Species. 2015.
18. Steyer K, Kraus RHS, Mölich T, Anders O, Cocchiario B, Frosch C, et al. Large-scale genetic census of an elusive carnivore, the European wildcat (*Felis s. silvestris*). *Conserv Genet*. 2016;17:1183–99.
19. Mueller SA, Reiners TE, Steyer K, von Thaden A, Tiesmeyer A, Nowak C. Revealing the origin of wildcat reappearance after presumed long-term absence. *Eur J Wildl Res*. 2020;66:94.
20. von Thaden A, Cocchiario B, Mueller SA, Reiners TE, Reinert K, Tuschscherer I, et al. Informing conservation strategies with museum genomics: long-term effects of past anthropogenic persecution on the elusive European wildcat. *Ecol Evol*. 2021;11:17932–51.
21. Mattucci F, Oliveira R, Lyons LA, Alves PC, Randi E. European wildcat populations are subdivided into five main biogeographic groups: consequences of Pleistocene climate changes or recent anthropogenic fragmentation? *Ecol Evol*. 2016;6:3–22.
22. Steyer K, Tiesmeyer A, Muñoz-Fuentes V, Nowak C. Low rates of hybridization between European wildcats and domestic cats in a human-dominated landscape. *Ecol Evol*. 2017;8 October:1–15.
23. Tiesmeyer A, Ramos L, Manuel Lucas J, Steyer K, Alves PC, Astaras C, et al. Range-wide patterns of human-mediated hybridisation in European wildcats. *Conserv Genet*. 2020;21:247–60.
24. Hartmann SA, Steyer K, Kraus RHS, Segelbacher G, Nowak C. Potential barriers to gene flow in the endangered European wildcat (*Felis silvestris*). *Conserv Genet*. 2013;14:413–26.
25. Pierpaoli M, Birò ZS, Herrmann M, Hupe K, Fernandes M, Ragni B, et al. Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Mol Ecol*. 2003;12:2585–98.
26. Gil-Sánchez JM, Jaramillo J, Barea-Azcón JM. Strong spatial segregation between wildcats and domestic cats may explain low hybridization rates on the Iberian Peninsula. *Zoology*. 2015;118:377–85.
27. Faure E, Kitchener A. An archaeological and historical review of the relationships between felids and people. *Anthrozoos A Multidiscip J Interact People Anim*. 2009;22:221–38.
28. Johansson F, Hüster H. Untersuchungen an Skelettresten von Katzen aus Haithabu (Ausgrabung 1966 – 1969). *Wachholtz; Neumünster*; 1987.
29. Ewing E *Fur in dress*. 168th ed. London: Batsford; 1981.
30. Beichman AC, Huerta-Sanchez E, Lohmueller KE. Using genomic data to infer historic population dynamics of nonmodel organisms. *Annu Rev Ecol Evol Syst*. 2018;49:433–56.
31. Pertoldi C, Randi E, Ruiz-González A, Vergeer P, Ouborg J. How can genomic tools contribute to the conservation of endangered organisms. *Int J Genomics*. 2016;2016:4712487.
32. Khan S, Nabi G, Ullah MW, Yousaf M, Manan S, Siddique R, et al. Overview on the role of advance genomics in conservation biology of endangered species. *Int J Genomics*. 2016;2016:3460416.
33. Mattucci F, Galaverni M, Lyons LA, Alves PC, Randi E, Velli E, et al. Genomic approaches to identify hybrids and estimate admixture times in European wildcat populations. *Sci Rep*. 2019;9:11612.
34. Meyer M, Spötl C, Mangini A. The demise of the Last Interglacial recorded in isotopically dated speleothems from the Alps. *Quat Sci Rev*. 2008;27:476–96.
35. Salonen JS, Helmens KF, Brendryen J, Kuosmanen N, Väiliranta M, Goring S, et al. Abrupt high-latitude climate events and decoupled seasonal trends during the Eemian. *Nat Commun*. 2018;9:2851.
36. Brauer A, Allen JRM, Mingram J, Dulski P, Wulf S, Huntley B. Evidence for last interglacial chronology and environmental change from Southern Europe. *Proc Natl Acad Sci*. 2007;104:450–5.
37. Croitor R, Brugal J-P. Ecological and evolutionary dynamics of the carnivore community in Europe during the last 3 million years. *Quat Int*. 2010;212:98–108.
38. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet*. 2017;49:303–9.
39. Lowe AJ, Boshier D, Ward M, Bacles CFE, Navarro C. Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity (Edinb)*. 2005;95:255–73.
40. McGrath MJ, Luyssaert S, Meyfroidt P, Kaplan JO, Bürgi M, Chen Y, et al. Reconstructing European forest management from 1600 to 2010. *Biogeosciences*. 2015;12:4291–316.
41. Klar N, Fernández N, Kramer-Schadt S, Herrmann M, Trinzen M, Büttner I, et al. Habitat selection models for European wildcat conservation. *Biol Conserv*. 2008;141:308–19.
42. Jerosch S, Kramer-Schadt S, Götz M, Roth M. The importance of small-scale structures in an agriculturally dominated landscape for the European wildcat (*Felis silvestris silvestris*) in central Europe and implications for its conservation. *J Nat Conserv*. 2018;41:88–96.
43. Lozano J, Virgós E, Malo AF, Huertas DL, Casanovas JG. Importance of scrub–pastureland mosaics for wild-living cats occurrence in a Mediterranean area: implications for the conservation of the wildcat (*Felis silvestris*). *Biodivers Conserv*. 2003;12:921–35.
44. Hensel M. Die gebräuchlichsten Jagdmethoden in der Zeit vom 16. bis zum 18. In: *Katalog Vom Jagen*. Hrsg. Museum Schloß Moritzburg; 1992.
45. Eckardt H. *Herrschaftliche Jagd, bäuerliche Not und bürgerliche Kritik: Zur Geschichte der fürstlichen und adligen Jagdprivilegien vornehmlich im südwestdeutschen Raum*. Göttingen: Vandenhoeck & Ruprecht; 1976.
46. Bacqué-Cazenave J, Bharatiya R, Barrière G, Delbecq J-P, Bouguiyoud N, Di Giovanni G, et al. Serotonin in animal cognition and behavior. *Int J Mol Sci*. 2020;21:1649.
47. Soubrié P. Reconciling the role of central serotonin neurons in human and animal behavior. *Behav Brain Sci*. 1986;9:319–35.
48. Leathwood PD. Tryptophan availability and serotonin synthesis. *Proc Nutr Soc*. 1987;46:143–56.
49. DeNapoli J, Dodman N, Shuster L, Rand W, Gross K. Effect of dietary protein content and tryptophan supplementation on dominance aggression, territorial aggression, and hyperactivity in dogs. *J Am Vet Med Assoc*. 2000;217:504–8.
50. van Hierden YM, de Boer SF, Koolhaas JM, Korte SM. The Control of Feather Pecking by Serotonin. *Behav Neurosci*. 2004;118:575–83.
51. Landsberg G, Milgram B, Mougeot I, Kelly S, de Rivera C. Therapeutic effects of an alpha-casozepine and L-tryptophan supplemented diet on fear and anxiety in the cat. *J Feline Med Surg*. 2016;19:594–602.
52. Goumas M, Lee VE, Boogert NJ, Kelley LA, Thornton A. The role of animal cognition in human-wildlife interactions. *Front Psychol*. 2020;11:3019.
53. Sih A, Ferrari MCO, Harris DJ. Evolution and behavioural responses to human-induced rapid environmental change. *Evol Appl*. 2011;4:367–87.
54. Hing S, Narayan EJ, Thompson RCA, Godfrey SS. The relationship between physiological stress and wildlife disease: consequences for health and conservation. *Wildl Res*. 2016;43:51–60.
55. Steyer K, Simon O, Kraus RHS, Haase P, Nowak C. Hair trapping with valerian-treated lure sticks as a tool for genetic wildcat monitoring in low-density habitats. *Eur J Wildl Res*. 2013;59:39–46.
56. Chapron G, Kaczensky P, Linnell J, von Arx M, Huber D, Andrén H, et al. Recovery of large carnivores in Europe's modern human-dominated landscapes. *Science*. 2014;346:1517–9.
57. Jarausch A, Harms V, Kluth G, Reinhardt I, Nowak C. How the west was won: genetic reconstruction of rapid wolf recolonization into Germany's anthropogenic landscapes. *Heredity (Edinb)*. 2021;127:92–106.
58. Barton N. The role of hybridization in evolution. *Mol Ecol*. 2001;10:551–68.

59. Mölich T, Vogel B. The wildcat as a target species for the reconnection of forest habitats: the long-term “A safety net for the wildcat” project. *Natur und Landschaft*. 2018;93:170–5.
60. Westekemper K, Tiesmeyer A, Steyer K, Nowak C, Signer J, Balkenhol N. Do all roads lead to resistance? State road density is the main impediment to gene flow in a flagship species inhabiting a severely fragmented anthropogenic landscape. *Ecol Evol*. 2021;11:8528–41.
61. Balzer S, Mölich T, Streif S, Tiesmeyer A, Thein J, Nowak C. Status der Wildkatze in Deutschland. *Natur und Landschaft*. 2018;93:146–52.
62. Andrews S. *FAstQC* a quality control tool for high throughput sequence data. 2010.
63. Ewels P, Magnusson M, Lundin S, Käller M. *MultiQC*: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32:3047–8.
64. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
65. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Consortium W. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet*. 2014;46:912–8.
66. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27:2987–93.
67. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–64.
68. Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. *pong*: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics*. 2016;32:2817–23.
69. Kassambara A, Mundt F. *factoextra*: extract and visualize the results of multivariate data analysis. R package version 1.0.7. 2017.
70. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011;475:493–6.
71. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
72. Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol Ecol*. 2016;25:1058–72.
73. Schiffels S, Wang K. MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent BT - Statistical Population Genomics. In: Duthel JY, editor. New York: Springer, US; 2020. p. 147–66.
74. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet*. 2021;108:1880–90.
75. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. *PLINK*: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
76. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation *PLINK*: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
77. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP data. *PLOS Genet*. 2013;9:e1003905.
78. Csilléry K, François O, Blum MGB. *abc*: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol*. 2012;3:475–9.
79. Meyermans R, Gorssen W, Buys N, Janssens S. How to study runs of homozygosity using *PLINK*? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*. 2020;21:94.
80. Weir BS, Cockerham CC. Estimating F-Statistics for the analysis of population structure. *Evolution (NY)*. 1984;38:1358–70.
81. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and *VCFTools*. *Bioinformatics*. 2011;27:2156–8.
82. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. *InterPro-Scan 5*: genome-scale protein function classification. *Bioinformatics*. 2014;30:1236–40.
83. Alexa A, Rahnenfuhrer J. *topGO*: enrichment analysis for gene ontology. R package version 2.42.0. 2020.
84. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57:289–300.
85. Kraysberg Y, Schwartz M, Brown TA, Ebralidse K, Kunz WS, Clayton DA, et al. Recombination of human mitochondrial DNA. *Science*. 2004;304:981.
86. Posada D, Crandall KA. Selecting the best-fit model of nucleotide substitution. *Syst Biol*. 2001;50:580–601.
87. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. *DnaSP 6*: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34:3299–302.
88. Katoh K, Rozewicki J, Yamada KD. *MAFFT* online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2019;20:1160–6.
89. Katoh K, Standley DM. *MAFFT* multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
90. Clement M, Snell Q, Walke P, Posada D, Crandall K. *TCS*: estimating gene genealogies. In: Proceedings 16th International Parallel and Distributed Processing Symposium. 2002. p. 7.
91. Leigh JW, Bryant D. *popart*: full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6:1110–6.
92. Martin SH, Davey JW, Jiggins CD. Evaluating the use of *ABBA-BABA* statistics to locate introgressed loci. *Mol Biol Evol*. 2015;32:244–57.
93. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the neandertal genome. *Science*. 2010;328:710–22.
94. Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. *Mol Biol Evol*. 2011;28:2239–52.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

