The genetic basis of seasonal coat colour variation in the least weasel, *Mustela nivalis*

Inês da Costa Miranda

Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia
Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-InBIO, Universidade do Porto)

2018

Orientador
José Melo-Ferreira, Investigador Auxilar, Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-InBIO, Universidade do Porto)

Coorientadora
Iwona Giska, Investigadora de Pós-Doutoramento, Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-InBIO, Universidade do Porto)

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Agradecimentos

Num abrir e fechar de olhos, dou por mim a terminar mais uma longa etapa, repleta de aprendizagens, crescimento e descobertas pessoais. Esta tese é o resultado de cinco anos de muito trabalho e não posso deixar de agradecer a todos os que, de uma forma ou de outra, contribuíram para que chegasse até aqui. Deixo, pois, o meu mais sincero obrigada:

Ao meu orientador, Dr. José Melo-Ferreira, por tudo e mais alguma coisa. Pela paciência e apoio incansáveis ao longo deste percurso, principalmente nestes últimos dias de exaustão. Pelas respostas sempre prontas e por todos os ensinamentos. Pelo exemplo. Pelo otimismo e entusiamo que me transmitiste desde o início. Por me motivares a fazer sempre mais e melhor. Por todas as oportunidades que me proporcionaste, por acreditares em mim e me confiares este projeto extraordinário. Não podia ter pedido melhor orientador.

To my co-supervisor, Dr. Iwona Giska, thank you for all the help and support during all the stages of this work. For the endless number of meetings to solve every little problem and for listening to all my doubts. For always being available to answer my questions and for sharing my interest for this system.

To Dr. Karol Zub, I thank for providing the samples from Poland included in this work, for his scientific input, and for all the previous work on this fascinating model. Also, for the wonderful pictures of weasels used in this thesis.

To Professor L. Scott Mills, I thank for the inspiring development of the study of seasonal coat colour change as a model to understand a relevant biological adaptation and the potential of evolutionary rescue facing climate change. Also, for kindly providing Figure 4, used in this thesis. I also thank Marketa Zimova for important discussions and for her help in sampling at the Swedish Museum of Natural History and the classification of winter coat colour in the specimens.

To the Institutions that kindly provided the samples used in this work, namely the Swedish Museum of Natural History (NRM, Stockholm, Sweden), the Zoological Research Museum Alexander Koenig (ZFMK, Bonn, Germany) and the Mammal Research Institute, Polish Academy of Sciences (Białowieża, Poland), and the curators and managers of the collections.

Aos restantes membros do grupo EVOCHANGE, por toda a ajuda incondicional. À Liliana, por teres feito as primeiras extrações e preparado as bibliotecas que foram essenciais neste projeto. Por aturares todas as mil e uma perguntas e me guiares durante o trabalho no laboratório. Ao Fernando, ao João e à Mafalda, pelas inúmeras opiniões e dicas e por estarem

sempre disponíveis para me ajudar com as análises. Por me escreverem os scripts que eu não teria conseguido sozinha e por facilitarem os meus primeiros passos neste mundo da bioinformática. Ao João, pela produção da pseudo-referência usada neste trabalho.

Ao Professor Paulo Célio Alves, por me ter dado a oportunidade de trabalhar neste meio fantástico, num dos momentos de maior indecisão que enfrentei no meu percurso académico. Pelo interesse por este projeto, por ter sempre uma palavra amiga e de encorajamento e por ter uma visão do mundo contagiante e enriquecedora. Se hoje aqui estou, muito se deve a si.

A todos os colegas e amigos que o CIBIO me deu, pelas conversas para descomprimir, pelas gargalhadas, pelos almoços e jantares e pelas viagens de metro. Em particular, ao Ivo, à Patrícia, à Sara e à Soraia, por terem partilhado tão de perto este último ano.

Aos meus colegas de Mestrado, por termos começado esta jornada juntos e, acima de tudo, por a terminarmos juntos. Pelos dias incontáveis entre cá e lá, pelas frustrações partilhadas e por me ajudarem a manter a sanidade mental. Em especial, à Catarina, por teres sido a minha parceira nesta aventura.

Ao melhor ano que Biologia já viu, por tornarem estes anos inesquecíveis. Por todas as memórias e por todos os obstáculos que passámos juntos. Porque ninguém melhor que vocês sabe o que significa chegar ao fim e ter a certeza que vos levo para a vida toda.

A todos os amigos que me fazem sorrir. Em particular, à Cat, pela ginástica de horários para encontros de última hora, pelas cusquices, planos, desabafos e por continuares do meu lado. À Rosa, por seres a companheira de todas as horas, por todo o apoio inesgotável neste ano de desencontros e por estares sempre à distância de uma chamada. À Vanessa, à Mafalda e à Maria, por todos os dias passados na faculdade, por me animarem e puxarem o pior de mim. Ao Mário, ao João e ao João, a família que eu escolhi, por tornarem sempre o meu dia melhor e por aceitarem as ausências.

Ao Luís, por estares lá incondicionalmente. Por tornares os dias maus em dias suportáveis e os dias bons em dias inesquecíveis. Por toda a força, coragem e paciência. Sem o teu apoio, este caminho teria sido muito mais difícil.

À minha família, por acreditarem em mim e por tornarem tudo mais fácil. Pelos sorrisos e abraços sempre disponíveis e por aturarem todos os meus devaneios. À Sofia, por me tirares a paciência dia sim, dia sim. À Filipa, por compreenderes o que fica por dizer e pelo teu humor retorcido tão compatível com o meu. Aos meus pais, por me darem asas para voar e por acreditarem nos meus sonhos.

Financial support

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through grants "CHANGE" PTDC/BIA-EVL/28124/2017 (Portuguese national funds), "2CHANGE" POCI-01-0145-FEDER-028124 (co-funded by European Regional Development Fund through Programa Operacional de Competitividade e Internacionalização – POCI-COMPETE – and national funds) and Investigador FCT IF/00033/2014/CP1256/CT0005 (co-funded by the European Social Fund through Programa Operacional de Potencial Humano – Quadro de Referência Estratégica Nacional and FCT). Sampling at the Swedish Museum of Natural History was supported by the SYNTHESYS program (grant SE-TAF-4695; EU FP7 agreement 226506).

Resumo

A sazonalidade ambiental impõe desafios importantes às espécies, através da variação das pressões seletivas ao longo do ano. Estas variações sazonais levaram à evolução de características sazonais (características fenológicas) que respondem a flutuações anuais do ambiente. A mudança sazonal da cor da pelagem é uma dessas adaptações. Mais de 20 espécies de mamíferos e aves mudam a cor da pelagem ou penas de castanho no verão para branco no inverno, o que permite aos indivíduos a manutenção da camuflagem durante todo o ano, em ambientes periodicamente cobertos por neve. No entanto, o valor adaptativo da cor branca de inverno está a ser severamente desafiado pelas alterações climáticas e consequente redução da extensão e persistência da cobertura de neve, causando desfasamentos entre a cor branca dos indivíduos e a cor dos habitats sem neve. Estes desfasamentos diminuem a aptidão dos indivíduos com cor branca de inverno e podem ameaçar a sobrevivências destas populações. No entanto, a existência de polimorfismos nestas espécies, com indivíduos que mantêm a cor castanha todo o ano, e, em particular, de populações polimórficas em que há coexistência de indivíduos com ambas as cores de inverno, pode ser uma fonte de rápida adaptação a climas em mudança. Perceber a base genética do polimorfismo da cor de inverno é, portanto, importante, não só para elucidar a evolução de uma notável adaptação sazonal, como também para perceber de que forma as espécies podem adaptar-se às alterações climáticas.

Neste trabalho, foi investigada a base genética da variação da cor sazonal na doninha-anã (*Mustela nivalis*), no contexto da estrutura populacional subjacente. Nesta espécie, existem dois tipos de coloração com distribuições geográficas distintas: i) o tipo *nivalis*, que muda para cor branca no inverno, apresenta uma linha de demarcação reta entre a pelagem dorsal e ventral e, geralmente, se distribui em regiões mais a norte, com mais cobertura de neve; e ii) o tipo *vulgaris*, que mantém uma cor castanha no inverno, tem uma linha de demarcação irregular e, geralmente, se distribui em latitudes mais a sul e/ou zonas costeiras, onde a cobertura de neve é mais reduzida. Na transição entre estas regiões, podem ser encontradas zonas polimórficas, onde coexistem indivíduos com diferentes cores de inverno, proporcionando uma oportunidade única para estudar a base genética do polimorfismo de inverno nesta espécie.

Espécimes dos dois tipos de coloração foram amostrados ao longo de uma zona polimórfica na Suécia, através da coleção do Museu Sueco de História Natural. Foi sequenciado o genoma completo de 40 indivíduos, com tamanho amostral equilibrado entre os tipos de coloração e com uma cobertura média aproximada de 1X por indivíduo. Análises

de estrutura populacional sugerem a existência de três populações geograficamente explícitas, com ocorrência de mistura genética. Foi encontrada uma correspondência parcial, mas não completa, com a distribuição dos tipos de coloração: foram identificadas duas populações com predominância de indivíduos *vulgaris*, no Este e Sul da região amostrada, e uma população a Norte, contendo maioritariamente indivíduos *nivalis*, mas também *vulgaris*. Análises filogenéticas sugerem que as populações a Este e Sul estão mais relacionadas. Estes resultados sugerem mistura genética entre as duas formas de cor, o que se reflete em valores de diferenciação médios baixos (F_{ST} = 0.06), estimados através de polimorfismos nucleotídicos simples (SNPs) distribuídos por todo o genoma. Foram conduzidos scans de associação ao longo de todo o genoma e três regiões genómicas, com diferenças extremas de frequências alélicas entre formas, foram identificadas como candidatas para associação à cor de inverno. O conteúdo genético destas regiões foi examinado, revelando dois genes com funções conhecidas na coloração, que ocorrem na mesma região candidata: *MC1R,* que está amplamente estabelecido como um dos genes de pigmentação mais importantes em mamíferos, e *SPIRE2*, cujo envolvimento no transporte de melanossomas já foi sugerido. Para explorar a relação entre fenótipos e genótipos individuais, 10 variantes desta região foram selecionadas, com especial cobertura no *MC1R*, e genotipadas por sequenciação em 83 espécimes da Suécia com fenótipo conhecido. Dois SNPs na sequência codificante do *MC1R* mostraram associação perfeita com o tipo de coloração, segregando de acordo com a relação de dominância sugerida por estudos anteriores: o alelo *vulgaris* dominante sobre o alelo *nivalis*. Estes SNPs candidatos foram ainda genotipados em amostras recolhidas ao longo de uma zona de transição na Polónia (da coleção da Academia Polaca de Ciências; N = 63) e em indivíduos resultantes de uma experiência de cruzamentos entre uma fêmea *nivalis* e um macho *vulgaris* (depositados no Museu de Investigação Zoológica Alexander Koenig; N = 33), confirmando a associação. Estas variantes resultam numa alteração amino-acídica que se inferiu que afete a função da proteína, tendo em conta a sua conservação a níveis filogenéticos mais profundos. Esta é, portanto, uma forte candidata a explicar o polimorfismo da cor de inverno, que poderá resultar de uma ativação constitutiva do recetor MC1R em indivíduos castanhos no inverno, resultando em menor responsividade do recetor ao seu antagonista, a proteína ASIP, e impedindo a muda de inverno nestes indivíduos. Por fim, as regiões genómicas candidatas foram testadas para detetar evidências de seleção positiva e os resultados sugerem a ocorrência de varrimentos seletivos em espécimes de inverno de cor castanha, mas não de cor branca. Isto poderá refletir um favorecimento dos alelos castanhos num período pós-glacial, caracterizado pelo aquecimento do clima e retração da cobertura de neve em direção a norte, reforçando o potencial para uma rápida adaptação às alterações climáticas.

No geral, este trabalho identificou o *MC1R* como o gene candidato mais forte para a determinação da variação da cor sazonal nas doninhas-anãs, reforçando as evidências (raras, mas existentes) de que o sistema *MC1R-ASIP* pode estar universalmente envolvido na determinação da coloração sazonal. No entanto, o impacto funcional das mutações identificadas deverá ser clarificado em investigação futura. Estão também lançadas as bases para perceber a evolução desta característica adaptativa nas doninhas-anãs, no género *Mustela* e nos mamíferos em geral. Para finalizar, este trabalho realça a importância das coleções de história natural como fontes de organismos localmente adaptados, relevantes para o estudo da base genética e da evolução de características adaptativas-chave em populações naturais.

Palavras-chave: polimorfismo da cor de inverno, mudança sazonal da cor da pelagem, *Mustela nivalis*, doninha-anã, sequenciação de genoma, estudo de associação, genómica aplicada a amostras de museu

Abstract

Environmental seasonality poses important challenges for species by varying the selective pressures along the year. These seasonal variations caused the evolution of seasonal traits (phenological traits), that respond to the yearly fluctuations of the environment. Seasonal coat colour change is one of such adaptations. Over 20 mammal and bird species change from summer-brown to winter-white coats, which allows individuals maintaining crypsis year-round, in environments periodically covered with snow. The adaptive value of white winter coats is, however, being severely challenged by climate change and the consequent decreases in snow cover extent and persistence, by causing mismatches between white coats and snowless backgrounds. These mismatches decrease the fitness of winter-white individuals and may threaten the survival of these populations. However, the existence of polymorphism in these species, with individuals maintaining the brown coat year-round, and, in particular, of polymorphic populations where specimens with winter-white and brown coats coexist, may be a source of rapid adaptation to the changing climate. Understanding the genetic basis of the winter coat colour polymorphism is thus important to shed light on the evolution of a striking seasonal adaptation and to understand how species can adapt to climate change.

In this work, the genetic basis of seasonal coat colour variation in the least weasel (*Mustela nivalis*) was investigated in the context of the underlying population structure. Two colouration types exist in the species, with distinct geographic distributions: i) the *nivalis* type, which moults to a white coat in winter, shows a straight demarcation line between the dorsal and ventral pelage, and is generally distributed in the northernmost latitudes with more snow cover, and ii) the *vulgaris* type, which maintains a brown coat in winter, has a ragged demarcation line, and is generally distributed in more southern latitudes and/or coastal regions where snow cover is more reduced. Polymorphic zones, where individuals with different winter colours coexist, can be found in the transition between these regions, providing a unique opportunity to study the genetic basis of winter colour polymorphism in the species.

Specimens of the two colour morphs were sampled across a polymorphic zone in Sweden, from the collection of the Swedish Museum of Natural History. The whole-genome of 40 individuals, with a balanced sample size between colouration types, was sequenced for an approximate mean coverage of 1X per specimen. Analyses of population structure suggested the existence of three geographically explicit genetic populations, with admixture. Some (but not complete) overlap with the distribution of the colour morphs was found: two predominantly *vulgaris* populations were identified in the East and South of the sampled range, together with one Northern population containing mostly *nivalis* specimens but also some *vulgaris*.

Phylogenetic analyses suggested the Eastern and Southern populations to be more closely related. These results suggest admixture between colour morphs, which is reflected in low mean differentiation values estimated from genome-wide SNPs (F_{ST} = 0.06). Whole-genome scans for differentiation were conducted, and three genomic regions with extreme allele frequency differences between both morphs were identified as candidates for association with the colour trait. The examination of the genetic content of these regions revealed two genes with known functions in colouration, occurring in the same candidate region: *MC1R*, widely established as one of the most important pigmentation genes in mammals, and *SPIRE2*, which has been suggested to be involved in the transport of melanosomes. To explore the link between the individual genotypes and phenotypes, 10 selected variants from this region, with particular coverage of *MC1R*, were genotyped by sequencing in 83 specimens from Sweden with scored phenotype. Two consecutive SNPs in the protein-coding sequence of *MC1R* were found to be perfectly associated with the colour morph and to segregate according to the expected dominance relationship suggested by previous studies, with the *vulgaris* allele being dominant over the recessive *nivalis* one. These candidate SNPs were further genotyped in samples collected across a coat colour transition zone in Poland (from the collection of the Polish Academy of Sciences; N = 63) and in one pedigree resulting from a crossing experiment between a *nivalis* female and a *vulgaris* male (deposited at the Zoological Research Museum Alexander Koenig; N = 33), confirming the association. These variants result in an amino-acid change that was predicted to affect protein function, given its conservation across deeper phylogenetic levels. This is, therefore, a strong candidate to underlie the trait polymorphism, which can be due to a constitutive activation of MC1R in winter-brown specimens that results in lower responsiveness to its antagonist, ASIP, hampering the winter moult. Finally, candidate genomic regions were tested for evidence of positive selection, and results suggest the occurrence of selective sweeps in winter-brown but not in winter-white specimens. This could reflect favouring of brown alleles in the post-glacial warming of the climate and snow cover retreat towards the north, reinforcing the potential for rapid adaptation to climate change.

Overall, this work identified *MC1R* as the strongest candidate to underlie the determination of seasonal coat colour variation in least weasels, adding to the rare but existent evidence that the *MC1R-ASIP* system may be universally involved in determining seasonal colouration. The functional impact of the identified mutations must, however, be clarified in future research. It also sets the foundation to understand the evolution of the trait in least weasels, in the *Mustela* genus, and in mammals in general. Finally, this work highlights the value of natural history collections as sources of locally adapted organisms to the study of the genetic basis and evolution of key adaptive traits in natural populations.

Keywords: winter coat colour polymorphism, seasonal coat colour change, *Mustela nivalis*, least weasel, whole-genome sequencing, genome-wide association study, museum genomics.

Index

List of Figures

Supplementary Figures

List of Tables

List of Abbreviations

AF – actin filament

Ala – alanine

ASIP – *agouti* signalling protein

bp – base-pair

cAMP – cyclic adenosine monophosphate

C – cysteine

CDS – coding sequence

CLR – composite likelihood ratio

D – aspartic acid

DCT – dopachrome tautomerase

DNA – deoxyribonucleic acid

DRG2 – developmentally regulated GTP-binding protein 2

EL – extracellular loop

Exo – exonuclease I

FANCA – Fanconi anaemia complementation group A

FMN1 – formin-1

FMN2 – formin-2

F_{ST} – fixation index

Glu/E – glutamic acid

GWAS – genome-wide association study

jSFS – joint site frequency spectrum

kb – kilo base-pair

Leu/L – leucine

Lys/K – lysine

LD – linkage disequilibrium

M – methionine

Mb – mega base-pair

MC1R – melanocortin-1 receptor

mtDNA – mitochondrial DNA

MYO5A – myosin-Va

MYO7A – myosin-VIIa

N – asparagine

NGS – next generation sequencing

NHM – natural history museum

NV – *nivalis* vs. *vulgaris*

P – proline

PCA – principal components analysis

PCR – polymerase chain reaction

PF – parental female

PZ – polymorphic zone

qPCR – quantitative polymerase chain reaction

QTL – quantitative trait loci

R – arginine

RNA – ribonucleic acid

SAF – site allele frequency

SAP – shrimp alkaline phosphatase

SFS – site frequency spectrum

SNP – single nucleotide polymorphism

SPIRE1 – spire-type actin nucleation factor 1

SPIRE2 – spire-type actin nucleation factor 2

SPNS2 – sphingolipid transporter 2

TCF25 – transcription factor 25

TMD – transmembrane domain

TYR – tyrosinase

TYRP1 – tyrosinase related protein 1

WB – winter-brown vs. winter-white

ZNF276 – zinc finger protein 276

α-MSH – α-melanocyte-stimulating hormone

1. Introduction

The environment imposes selective pressures on populations that determine the fitness of individuals. Favourable characteristics, i.e. adaptive traits, tend to increase in frequency in populations across generations, allowing them to better persist in that particular environment. These characteristics can include morphological, physiological, or behavioural traits, which enhance both their survival and reproductive success in a given habitat (Bijlsma and Loeschcke, 2005). Seasonality in key adaptive traits is widespread in nature and is crucial for maintaining species' fitness because it allows coping with environmental seasonality (Varpe, 2017). Seasonally adapted traits include, for example, reproduction, migration, hibernation, or crypsis. The phenology of these events, i.e. their timing along the year, is best explained by a complex interaction between biotic and abiotic factors, particularly, the genetic makeup of the organisms, which allows them to react to photoperiod, temperature, precipitation, or other environmental factors (Forrest and Miller-Rushing, 2010; Varpe, 2017).

The evolution of many adaptive traits in response to selective pressures is well documented (e.g. Kingsolver *et al.*, 2001; Grant and Grant, 2006; Elzinga *et al.*, 2007). Nonetheless, climate change is leading to rapid modifications in environmental conditions, such as increasing temperatures and warming oceans (Solomon *et al.*, 2009). These quick changes impose new conditions that may compromise the persistence of species if populations are unable to rapidly adapt (e.g. Pauli *et al.*, 2013; Zimova *et al.*, 2016). Rapid adaptation may thus be essential for species that are unable to disperse to more favourable habitats or that have limited plasticity of key phenotypes (Hoffmann and Sgrò, 2011). Interestingly, some studies have reported polymorphisms in phenological traits, such as egg hatching in the green oak tortix (Du Merle, 1999), breeding dates in barn swallows (Caprioli *et al.*, 2012), or winter colour in mammal and bird species (Jones *et al.*, 2018; Mills *et al.*, 2018) that are explained by genetic factors. These polymorphisms can facilitate microevolution of these traits (Caprioli *et al.*, 2012; Zimova *et al.*, 2018), thus facilitating the evolutionary rescue of species challenged by global warming (Mills *et al.*, 2018). Hence, it is important to improve our understanding of the genetic architecture of polymorphic phenological traits and how this variation can be crucial for the adaptation of organisms to changing climates. Seasonal coat colour change is one of the most striking examples of seasonal adaptations.

1.1. Seasonal coat colour change

1.1.1. The importance of colouration as an adaptive trait

Colouration is one of the most remarkable traits in natural populations, which has strong impacts on multiple key mechanisms that influence both reproduction and survival. As a consequence, numerous colour patterns are observed in nature. From insects to mammals, colouration has evolved, allowing populations to adapt to the environment (see review in Protas and Patel, 2008) and being associated with different functions, as thermoregulation, communication, or sexual selection (Booth, 1990; Roulin, 2004). For example, different colour polymorphisms have been implicated in assortative mating, such as head colour in the Gouldian finch (*Erythrura gouldiae*) (Pryke and Griffith, 2007) and body colour in the Midas cichlid species complex (Barlow *et al.*, 1977; Barlow, 1992). Likewise, the wing colouration in the dragonfly *Zenithoptera lanei* has an important role in intraspecific communication, functioning as a cue for rivals' recognition in territorial fights (Guillermo-Ferreira *et al.*, 2015).

Colouration is also particularly important in predator-prey interactions, and important colour adaptations include aposematism, mimicry, and crypsis (Booth, 1990). Aposematism occurs when a species displays vivid colours that suggest toxicity or distastefulness, leading predators to avoid them, as seen in the poison frog family Dendrobatidae (Summers and Clough, 2001) and in numerous insects [e.g. the Asian ladybird beetle, *Harmonia axyridis* (Bezzerides *et al.*, 2007) and the butterfly genus *Heliconius* (Parchem *et al.*, 2007)]. Mimicry occurs when a prey resembles another species to avoid selection from the predator, such as some harmless species of snakes that display the same strong colour patterns of New World venomous coral snakes (Rabosky *et al.*, 2016). Crypsis is the mechanism that allows concealment through matching of the colour of the body to the background. Unlike aposematism and mimicry, which make species conspicuous in their surroundings, crypsis allows animals to remain camouflaged and can either increase predation rates (for predators) or decrease predation risk (for prey). Three striking examples of adaptive evolution that favour concealment are the polymorphisms described in the peppered moth (*Biston betularia*) (Bishop, 1972), the rock pocket mice (*Chaetodipus intermedius*) (Nachman *et al.*, 2003), and the beach mice (*Peromyscus polionotus*) (Hoekstra *et al.*, 2006). In all these cases, when animals were exposed to environments with contrasting colours, distinct body colours evolved in the species, allowing concealment on distinct backgrounds. Although body colouration is usually maintained through the life of an animal, some species are able to quickly change their colour patterns in response to abrupt changes in the background colour or to different predators with different visual capacities (Stuart-Fox and Moussalli, 2009). Additionally, some species change

colour seasonally, depending on circumstances like environmental changes and the breeding period (Booth, 1990).

1.1.2. Seasonal colour adaptations

Seasonal coat colour change is a particularly striking colour adaptation that allows the maintenance of camouflage year-round. It consists of at least two yearly plumage or pelage moults that allow animals to maintain fitness through annual environmental modifications. In general, during the autumn moult, the summer-brown coat is shed and replaced by a winter-white one; in spring, the opposite process occurs (reviewed in Zimova *et al.*, 2018). This allows animals to match the presence or absence of snow on the ground, with a clear adaptive role in assuring concealment according to the predominant background colour (Mills *et al.*, 2018). Although shedding and regrowth of the coat is part of the hair (Stenn and Paus, 2001) and feather (Chen *et al.*, 2015) growth cycles, seasonal colour changes from summer-brown to winter-white are restricted to some species that inhabit arctic or temperate regions in the Northern Hemisphere. These include, for example, the rock ptarmigan (*Lagopus muta*) (Vaurie, 1965), the arctic fox (*Vulpes lagopus*) (Våge *et al.*, 2005), the collared lemming (*Dicrostonyx groenlandicus*) (Gower *et al.*, 1993), the Siberian hamster (*Phodopus sungorus*) (Logan and Weatherhead, 1978), the mountain hare (*Lepus timidus*) (Angerbjörn and Flux, 1995), the snowshoe hare (*L. americanus*) (Severaid, 1945), the long-tailed weasel (*Mustela frenata*) (Bissonnette and Bailey, 1944), and the stoat (*M. erminea*) (Rust, 1965) (reviewed in Zimova *et al.*, 2018).

The moult cycle, like other seasonal traits, is predominantly dependent on the circannual rhythm controlled by photoperiod. In mammals, the process is mediated by the plasma levels of two hormones: melatonin and prolactin (Hofman, 2004). The photoperiodic information is collected in the retina and transmitted to the suprachiasmatic nucleus of the hypothalamus, from where a signal is forwarded to the pineal gland, where melatonin is produced (Moore, 1995). This production occurs during the night, thus levels of melatonin in the plasma are inversely proportional to day length (Goldman, 2001). Therefore, short day lengths originate higher levels of melatonin, which further inhibit the production of prolactin by the pituitary gland (Lincoln *et al.*, 2006), triggering the autumn moult. During the spring moult, the opposite process occurs – long day lengths result in low levels of melatonin, which trigger an increase in the level of prolactin that induces the spring moult. However, the exact mechanism through which prolactin regulates the moult is yet poorly understood. Several studies have been conducted in which moults were artificially induced through changes of the melatonin-prolactin

balance (e.g. Rust, 1965; Duncan and Goldman, 1984; Gower *et al.*, 1993). On the one hand, for example, the administration of prolactin to Siberian hamsters with the winter coat stimulated the production of pigment, inducing changes to a brown pelage, despite being kept in short light periods (Duncan and Goldman, 1984). On the other hand, the artificial reduction of prolactin levels induced the moult to winter-white in collared lemmings exposed to a long day period (Gower *et al.*, 1993) and in stoats undergoing the spring moult (Rust and Meyer, 1969). Similarly, the surgical removal of the pituitary gland (and consequent inhibition of prolactin production) induced the growth of white fur on stoats displaying both winter-white and summer-brown colours (Rust, 1965). Furthermore, Bissonnette and Bailey (1944) induced both autumn and spring moults in the long-tailed weasel by altering the photoperiod.

1.1.3. The cost of mismatch and the value of winter coat colour polymorphism

Climate change is causing decreases in both snow cover duration and extent across the Northern Hemisphere (Brown and Mote, 2009; Pederson *et al.*, 2011), likely related with seasonal changes in the cycle of tropospheric temperature (Santer *et al.*, 2018). Importantly, projections for the future predict further snow decreases (Danco *et al.*, 2016; Musselman *et al.*, 2017). Because the moult cycle is mainly triggered by photoperiod (despite some influence of other environmental parameters, as temperature, in the rate of change in some species; e.g. Rothschild, 1942; Rust, 1962; Hewson, 1973), reductions in snow cover cause mismatches between the coat and background colours when individuals change to a white coat in winter, hence challenging the adaptive value of seasonal coat colour change. For example, a study conducted in the snowshoe hare (*Lepus americanus*) showed that the disruption of crypsis decreases the survival rate of the species, due to increased number of predator attacks (Zimova *et al.*, 2016). Similar results were obtained when exposing models of the least weasel (*Mustela nivalis*) with both brown and white colour to contrasting backgrounds (Atmeh *et al.*, 2018). Furthermore, seasonal coat colour mismatches have already been associated to range contractions in the snowshoe hare (Sultaire *et al.*, 2016) and the rock ptarmigan (*Lagopus muta*) (Imperio *et al.*, 2013). These examples highlight the importance of understanding which type of adaptations can be developed to face snowless backgrounds. For example, behavioural responses have been investigated in different species. Studies in ptarmigans have shown that male rock ptarmigans in Canada soiled their plumage after snowmelt to reduce conspicuousness (Montgomerie *et al.*, 2001), and willow ptarmigans (*L. lagopus*) in Norway selected feeding areas that matched their colour during spring moult, even if these areas offered low-quality food (Steen *et al.*, 1992). Yet, no behavioural changes were detected in snowshoe hares when exposed to contrasting backgrounds (Zimova *et al.*, 2014).

In addition, the plasticity of the moult phenology has also been studied. Studies on the snowshoe hare showed that the initial dates of the moults and the rate of the autumn moult remained unaltered under different environmental conditions (Mills *et al.*, 2013), suggesting a very limited effect of plasticity in the initiation of the moult. Some plasticity was, however, inferred in the rate of the spring moult (slower in years with longer snow cover; Mills *et al.*, 2013; Zimova *et al.*, 2014), which was also observed in the mountain hare (*Lepus timidus*) (Flux, 1970). Moreover, comparisons of the moulting patterns of the least weasel over about 40 years suggested very limited plasticity (Atmeh *et al.*, 2018). Overall, behavioural and phenotypic plasticity may be insufficient to help seasonal coat colour changing species coping with climate-induced mismatch. However, the existence of winter coat colour polymorphism in many species that seasonally change colour (i.e. the presence of both winter-white and winter-brown individuals) may be a key aspect to allow rapid adaptation of these species to global warming (Mills *et al.*, 2018).

Although seasonal coat colour changing species tend to be winter-white in an important part of their distribution ranges, variations in winter colour exist, often discrete, due to either incomplete moults or the expression of a distinct winter coat colour. For example, in rock ptarmigans from the Amchitka Island (*Lagopus mutus gabrielsoni*), females have partially dark feathers covering the head, and both sexes (though mostly females) have dark feathers in the upper parts, as part of the usual winter plumage (Jacobsen *et al.*, 1983). Also, specimens of the least weasel from Missouri and Iowa (USA) were found with a winter pelage consisting of white fur with brown spots and patches or with a mid-dorsal brown stripe (Easterla, 1970). These populations tend to occur in areas with patchy or sporadic snow cover (Jacobsen *et al.*, 1983; Zimova *et al.*, 2018), thus assuring concealment against inconsistent background colour. Notably, in some populations, individuals express a non-white winter coat. For example, mountain hares introduced to the Faroese Islands initially changed to white in winter but, over the generations, progressively started changing to a grey winter coat (Degerbøl, 1940). Other populations express the brown coat year-round, as willow ptarmigans from Scotland (*Lagopus lagopus scotica*) (Skoglund and Hoglund, 2010), mountain hares from Ireland (*Lepus timidus hibernicus*) (Flux and Angermann, 1990), snowshoe hares in the Pacific Northwest coast of North America (Nagorsen, 1983), or least weasels from England (King, 1979) and Southern Europe (King and Powell, 2007). Translocation and common garden studies in many of the species where winter-white and brown specimens exist strongly suggest that the polymorphism is genetically determined (see Zimova *et al.*, 2018 and references therein). Importantly, some populations are polymorphic for winter coat colour, with winter-white and winter-brown individuals coexisting. Mills *et al.* (2018) studied the distribution of distinct winter colours (white or brown) in eight colour-changing species and showed that these species

display clinal variation of winter colour that can be explained by snow cover variables. Higher probability of being winter-white was found to be associated with more persistent snow cover, while in regions with rare or inexistent snow there is higher probability of being winter-brown. Moreover, their predictions suggest that regions where the duration of snowpack is short or inconsistent are associated with polymorphic populations where both winter-white and winter-brown morphs co-occur. In addition, these results suggest that phenotypic variation in winter colours across environmental gradients can be explained by local selection for maintenance of camouflage against distinct backgrounds. These conclusions are consistent with previous reports of populations with polymorphism in winter colour, such as stoats from Scotland and Northeast England, where higher percentage of winter-white individuals was associated to a higher number of days with snow cover (Hewson and Watson, 1979).

In a context of climate change, winter phenotypes are expected to undergo strong local selective pressures according to snow cover duration, leading to microevolution of winter colour (Zimova *et al.*, 2018). Polymorphic zones where different winter colours co-exist should, therefore, be sources of evolutionary rescue (Mills *et al.*, 2018), i.e. where standing variation may allow rapid adaptation.

1.2. The genetic basis of colour traits

Numerous studies have been conducted that explored the physiology and genetics of pigmentation in vertebrates (reviewed in Hubbard *et al.*, 2010). Colouration in both mammals and birds is strongly determined by the production of melanin pigments in specialised cells, the melanocytes. Melanin synthesis occurs in melanosomes, organelles that reside in melanocytes in variable number, size, and density (Hearing and Tsukamoto, 1991). Two types of pigment can be produced: eumelanin (black to brown colour) and pheomelanin (red to yellow colour). The pigment is transported to keratinocytes, during the development of hairs or feathers, where it is deposited (Slominski *et al.*, 2004). The type of melanin produced is primarily determined by interactions between the melanocortin-1 receptor (MC1R), present in the membrane of melanocytes, and two ligands: α-melanocyte-stimulating hormone (α-MSH) and *agouti* signalling protein (ASIP) (see Figure 1). α-MSH binds to MC1R and functions as an agonist (Rouzaud *et al.*, 2003), increasing its activity. Consequently, MC1R activates adenylyl cyclase signalling pathway, leading to increased levels of cAMP that ultimately result in increased activity of tyrosinase (TYR), a key enzyme in the melanogenesis pathway (Hearing and Tsukamoto, 1991), thus inducing the eumelanin synthesis. In contrast, ASIP is an antagonist of α-MSH (Lu *et al.*, 1994). When it binds to MC1R, the receptor is inhibited, TYR

activity is reduced, and the pigment synthesis is switched to pheomelanin (Barsh, 1996). Additionally, ASIP has been shown to inhibit the differentiation of melanoblasts into melanocytes (Aberdam *et al.*, 1998; Sviderskaya *et al.*, 2001), hence decreasing melanocytes' biological functions and inhibiting melanogenesis, which can result in the complete absence of pigments and white coat colour (Le Pape *et al.*, 2009).

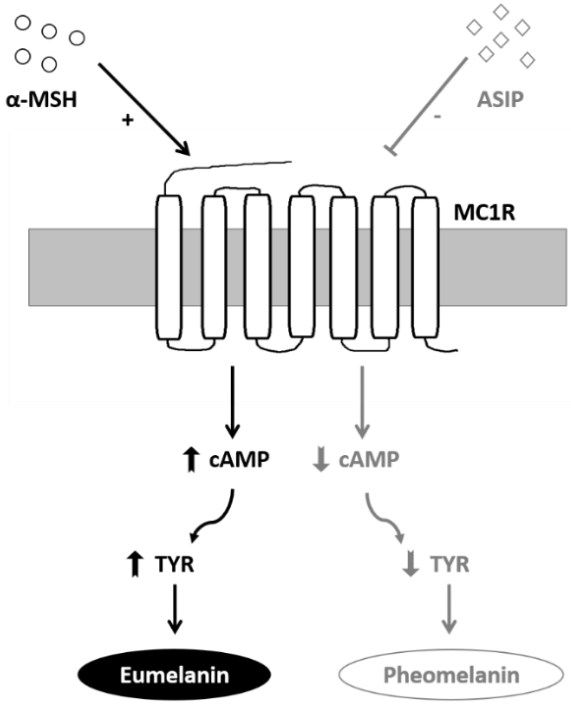

Figure 1 – Schematic representation of the regulation of melanocortin-1 receptor (MC1R) activity through α-melanocyte-stimulating hormone (α-MSH) and *agouti* signalling protein (ASIP). While α-MSH stimulates the synthesis of eumelanin through activation of a cAMP-dependent pathway and increased tyrosinase (TYR) activity, ASIP inhibition leads to the synthesis of pheomelanin. Adapted from Rouzaud *et al.* (2003) and Slominski *et al.* (2004).

An important number of genes involved in pigmentation has been identified in numerous studies using both natural populations and domestic animals. For example, in mammals, mice have been the most widely studied model, providing important knowledge on the regulation of this trait and leading to the identification of a substantial number of candidates that explain distinct colouration patterns (Bennett and Lamoreux, 2003; Hoekstra, 2006). Mutations in *MC1R* and *ASIP* genes seem to be particularly recurrent in explaining permanent colour variation in distinct species [e.g. dog (Everts *et al.*, 2000), rabbit (Fontanesi *et al.*, 2006), or leopard (Schneider *et al.*, 2012)]. However, when it comes to seasonal coat colour change, little is known about its genetic basis. Experiments conducted with the Siberian hamster (*Phodopus sungorus*) explored tyrosinase and melanin levels in hair follicles during moult periods, and results have shown that tyrosinase levels presented a peak during both spring

and autumn moults, whereas melanin levels showed a peak only during spring moult, when the fur turns brown (Logan and Weatherhead, 1978). These findings suggested the interruption of the melanogenesis pathway somewhere between tyrosine production and melanin synthesis during the autumn moult, possibly due to regulatory mechanisms. A recent work conducted by Ferreira *et al.* (2017) identified numerous genes differentially expressed between spring moult stages in the snowshoe hare, including genes involved in circadian rhythms, hair follicle morphogenesis, and melanogenesis. Interestingly, the list included three well-established pigmentation genes: *ASIP*, *MYO7A*, and *SPNS2*. These results provided the first picture of major transcriptional changes occurring during the moult, which can guide new studies to further explore the mechanisms underlying this trait.

Although the genetic basis of seasonal coat colour change is still poorly studied in most of the species, some works have identified genomic regions involved in intraspecific winter coat colour variation. For instance, the Arctic fox (*Vulpes lagopus*) presents two colour morphs, white and blue, which display distinct summer and winter coat. Whereas the first presents a summer dark coat with light ventral colour and a completely white winter coat, the second displays a uniform dark grey summer coat and does not change to winter-white. Two non-synonymous mutations leading to two aminoacidic substitutions that result in the blue variant were identified in *MC1R* gene (Våge *et al.*, 2005). Nonetheless, other works have failed to determine the genetic basis of winter colour variation in other species. Nunome *et al.* (2014) explored genetic variation in three candidate genes (*MC1R*, *ASIP*, and *TYR*) in the Japanese hare (*Lepus brachyurus*) but found no differences between winter-white and winter-brown individuals. Likewise, Skoglund and Hoglund (2010) targeted the coding regions of four genes involved in melanin pigmentation (*MC1R*, *TYR*, *TYRP1*, and *DCT*) but found no polymorphisms associated with winter colour differences in willow ptarmigans. On the contrary, Jones *et al.* (2018), using whole-genome scans, have recently shown that genetic variation in the *ASIP* regulatory region is associated with winter colour variation in the snowshoe hare. This work showed *cis*-regulatory differences in *ASIP* expression during the autumn moult, with overexpression of *ASIP* in the winter-white allele when compared to the winter-brown allele. Additionally, it showed that the winter-brown variant of the gene was acquired from a neighbouring winter-brown species, the black-tailed jackrabbit (*Lepus californicus*), through hybridisation and adaptive introgression.

1.3. Seasonal coat colour in *Mustela nivalis*

1.3.1. *Mustela* spp. and seasonal coat colour change

Genus *Mustela* is the largest from the Mustelidae family, and the number of species described ranges from 13 to 17, according to different authors (e.g. Corbet, 1978; Corbet and Hill, 1991; Abramov, 2000). Commonly, these animals are classified as weasels, ferrets or polecats, and minks, depending on both size and colour of the body. Generally, weasels are the smallest and present brown back and white to yellow belly, ferrets are bigger than weasels and are characterised by dark masks on the face and black legs and tails, and minks display a completely brown body, although they may present lighter patches in the chin and/or chest (King and Powell, 2007).

Three of *Mustela* species change their coat colour seasonally from summer-brown to winter-white: the long-tailed weasel (*M. frenata*), the stoat (*M. erminea*), and the least weasel (*M. nivalis*) (see e.g. King and Powell, 2007). All three have populations that maintain the brown coat year-round, usually in the southernmost parts of their ranges, with clinal variation of the winter colour phenotype (Mills *et al.*, 2018). Several studies explored the phylogenetic relationships of *Mustela* species based on mitochondrial DNA (Kurose *et al.*, 2000; Sato *et al.*, 2003; Kurose *et al.*, 2008), nuclear DNA (Sato *et al.*, 2003), or both (Koepfli *et al.*, 2008; Sato *et al.*, 2012). These phylogenies show that the brown-to-white seasonal coat colour change is scattered throughout the phylogeny (Figure 2), questioning how repeated the evolution of seasonal coat colour change and of winter colour polymorphism are.

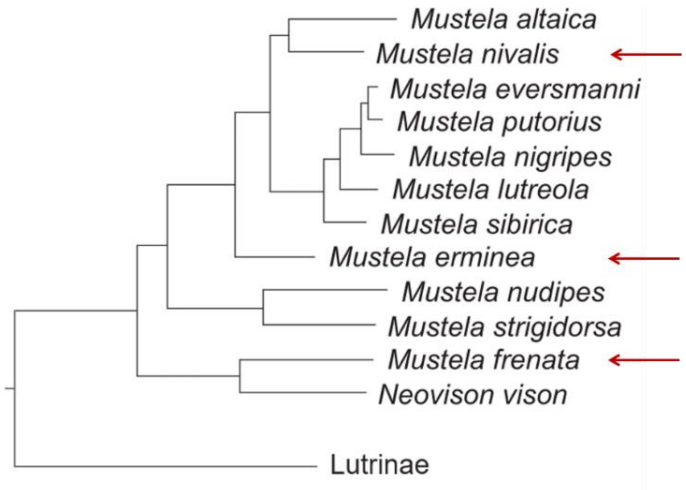

Figure 2 – Phylogeny of the Mustelinae (*Mustela* and *Neovision* spp.). Species that seasonally change colour from brown to white are indicated with an arrow. Adapted from Koepfli *et al.* (2008).

1.3.2. Colour patterns and seasonal colour change in the least weasel

The least weasel (*Mustela nivalis*) is the smallest species from the Mustelidae family and it is widely distributed in the Holarctic, occupying most of Europe, Asia, and North America (Sheffield and King, 1994). Furthermore, it was introduced in some islands, including Malta, Crete, and Azores (Corbet, 1978). Numerous subspecies have been described, and the accepted number differs between authors (e.g. Sheffield and King, 1994; Abramov and Baryshnikov, 2000; King and Powell, 2007). Two subspecies consistently recognised are *M. n. nivalis* and *M. n. vulgaris*. According to the distribution ranges described by Abramov and Baryshnikov (2000), *nivalis* occurs in Northern Europe (particularly Scandinavia and the northernmost part of Eastern Europe) and most of Asia, whereas *vulgaris* occurs in Central and Southern Europe. These authors further recognise a third European subspecies, *M. n. boccamela*, distributed through the Southwest, including most Iberian Peninsula, Southern France, and the Italian Peninsula. Yet, King and Powell (2007) do not consider the existence of *boccamela* and include those specimens in the *vulgaris* subspecies, following what was previously proposed by van Zyll de Jong (1992). Current classifications are, however, based exclusively in morphological and phenotypic data.

M. n. nivalis and *M. n. vulgaris* are usually distinguished by three main characteristics. First, concerning body size, *nivalis* is smaller than *vulgaris*, following the global pattern of decreasing body size towards northern regions seen within the species (Abramov and Baryshnikov, 2000; King and Powell, 2007). Second, regarding the summer coat, *nivalis* is characterised by a straight demarcation line between the brown back and the white belly. On the contrary, *vulgaris* exhibits a ragged line, often with variable dark spots or patches among the white venter and two spots behind the angle of the mouth called gular spots (Sheffield and King, 1994; King and Powell, 2007). Moreover, the details of *vulgaris* demarcation line are individually unique and kept through moults, allowing the identification of each specimen (Linn and Day, 1966). These colour patterns are designated, according to Frank (1985), as colouration type I and type II, respectively (Figure 3A-C). Third, while *nivalis* seasonally changes coat colour from summer-brown to winter-white, *vulgaris* maintains the brown pelage year-round (Figure 3D; King and Powell, 2007). The two colouration characteristics, demarcation line and winter colour, have been suggested to be tightly associated (*nivalis* – straight line and winter-white; *vulgaris* – ragged line and winter-brown). However, a few observations suggest a possible dissociation of the two colour traits. King and Powell (2007) described the occurrence of Mediterranean populations, which remain brown in winter, with colouration type I. On the contrary, some individuals with colouration type II have been reported to display winter-white coat, in regions from Central or Southern Europe (e.g. Cavazza, 1909; Abramov and Baryshnikov, 2000; Zima

and Cenevova, 2002). Additionally, these colouration types are not exclusive to *nivalis* and *vulgaris* but rather shared with other subspecies described within *M. nivalis* (Abramov and Baryshnikov, 2000). Hereafter, *nivalis* and *vulgaris* will designate colouration types, rather than subspecies.

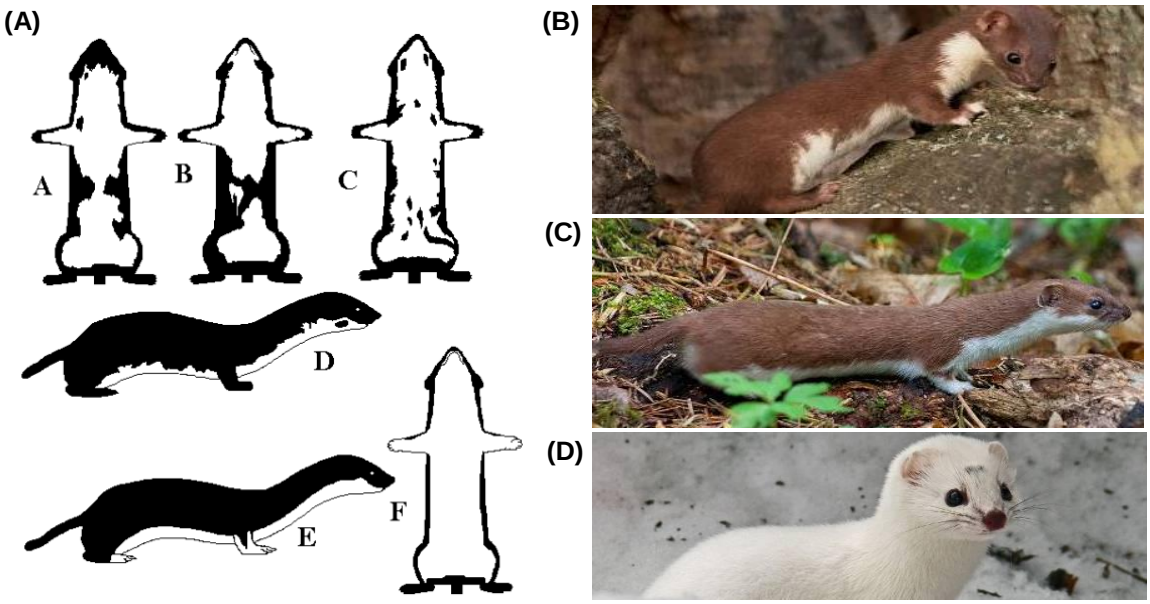

Figure 3 – Coat colour in the least weasel, *Mustela nivalis*. (A) Schematic representation of summer colour type II (A, B, C, and D) and type I (E and F). Adapted from Abramov and Baryshnikov (2000). Specimens of (B) *vulgaris* in the summer coat, (C) *nivalis* in the summer coat, and (D) *nivalis* in the winter coat. Photo credits: Karol Zub.

Least weasels undergo two moults per year, as most mammal species. In western Europe and the United Kingdom, the winter pelage is worn from November to March, whereas the summer pelage occurs from May to September. While the autumn moult begins around September or October and ends in November, the spring moult goes from March until June (King, 1979). However, in Poland, specimens undergoing spring moult were detected as soon as early February (Atmeh *et al.*, 2018) and late autumn moults were occasionally observed (Zub, personal communication). Records from the Polish population of *M. nivalis* show that the number of *nivalis* specimens in the Białowieża Forest, where *nivalis* and *vulgaris* co-exist, decreases proportionally to the duration of snow cover in winter. This is probably caused by increased predation rate of *nivalis* individuals, due to strong colour contrast against the surroundings, suggesting a strong fitness cost of mismatch (Atmeh *et al.*, 2018).

Regarding the genetic basis of these traits, little is yet known. Frank (1985) conducted crossing experiments between a *nivalis* female collected in Sweden and a *vulgaris* male from Germany and parental backcrosses, that shed some light on the genetic patterns of inheritance

of both summer colour types and winter-whitening. His results supported the dominance of type II (ragged demarcation line) over type I (straight line). Moreover, the maintenance of the winter-brown phenotype showed complete association to type II, being dominant over the moulting to winter-white, associated with type I.

Phylogeographic studies conducted in the least weasel, though mainly based on mitochondrial DNA (Lebarbenchon *et al.*, 2010; McDevitt *et al.*, 2012; Rodrigues, 2015), suggest no coincidence between population structure and the distribution of both winter colour morphs, both at global (Lebarbenchon *et al.*, 2010) or fine scales (McDevitt *et al.*, 2012). This suggests a role of local selective pressures, rather than historical divergence between distinct genetic populations, in explaining the distribution of winter coat colour. Transitions of distinct winter colour morphs occur in different areas in Europe, originating polymorphic populations where discrete variation of winter coat colour exists (Figure 4; Mills *et al.*, 2018). One particularly narrow transition zone, with about 100 km, occurs in southern Sweden, from Noorköping to Uppsala. It marks the boundary between populations of *nivalis* and *vulgaris* that not only have distinct winter coats but also conserve distinct summer patterns (Stolt, 1979 as cited in King and Powell, 2007). Because the two morphs are completely interfertile (Frank, 1985), gene flow may occur in this region, as well as in other polymorphic areas. Therefore, these transition zones can provide a unique opportunity to dissect the genetic mechanisms that underlie differences in winter coat colour in *Mustela nivalis*.

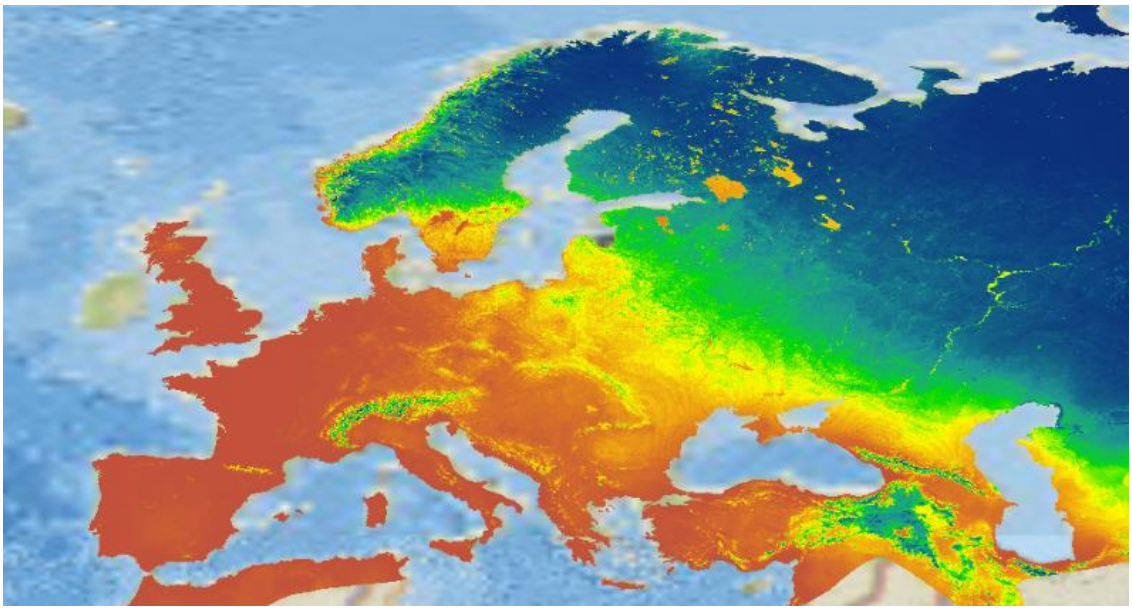

Figure 4 – Clinal variation of winter coat colours of *Mustela nivalis*. Warm colours (red/orange) represent higher probability of winter-brown phenotype, whereas cold colours (blue) represent higher probability of winter-white phenotype. Polymorphic areas are represented in yellow/green. Adapted from Mills *et al.* (2018).

1.4. Genotype-to-phenotype associations and signatures of selection

The evolution of adaptive characteristics in new environments is often explained by specific genetic variants that determine their phenotypic expression. These variants can arise from *de novo* mutations, i.e. they are newly formed, or from standing genetic variation, i.e. they were previously present in the population and were favoured by new selective forces, resulting in a rapid increase in frequency (Barrett and Schluter, 2008). New adaptive genetic variation may be also acquired from a closely related species, through hybridisation (e.g. Jones *et al.*, 2018; Seixas *et al.*, 2018). Variants underlying adaptive phenotypes can range from single-nucleotide polymorphisms (Nachman *et al.*, 2003) to more complex structural variants [e.g. deletions (Liu *et al.*, 2017) or insertions (Speed *et al.*, 2015)]. Moreover, they can occur at different genomic levels, affecting protein-coding genes in various manners. For example, mutations in protein-coding sequences may affect the function of the protein, as was shown in the determination of the coat colour of rock pocket mice (Nachman *et al.*, 2003). Alternatively, changes in non-coding regulatory regions may result in changed levels of gene expression of a certain trait, as seen in the determination of distinct winter colours in the snowshoe hare (Jones *et al.*, 2018).

Hence, to understand phenotypic differences in adaptive traits, such as colouration, different methods have been developed that can help establish links between the expressed phenotype and the underlying genotype(s). One of the traditionally applied strategies is the candidate gene approach, which relies on solid *a priori* knowledge of the biochemical or functional pathways that directly or indirectly regulate the trait of interest, from which candidate genes can be drawn and studied. However, this technique is necessarily limited by the existence of background information, the subjectivity inherent to the process of choosing which candidates to explore, and difficulties in selecting all possible candidates (Zhu and Zhao, 2007). Methods developed in the framework of next-generation sequencing (NGS) overcome these difficulties. The development of NGS allowed the transition from analyses comprising only a very small proportion of the genome to studies that comprehensively analyse information from the entire genome (Metzker, 2010; Mardis, 2011). Massive parallel sequencing has been used to obtain unprecedented amounts of genomic information using different sequencing strategies, as whole-genome sequencing (e.g. Cao *et al.*, 2011), target enrichment approaches that capture specific genomic regions [e.g. exomes (Ng *et al.*, 2009) or mitochondria (Briggs *et al.*, 2009)], or RNA-sequencing (e.g. Ferreira *et al.*, 2017). The use of NGS allows the simultaneous study of thousands of markers, increasing the resolution and power to answer many different genetic and evolutionary questions (Davey *et al.*, 2011). Additionally, the amounts and types of sequencing data that can now be generated allow analysing non-model organisms without previously available resources (Stapley *et al.*, 2010),

even though the existence of a close reference genome is beneficial. Genome-wide data allow scanning the genome for patterns indicating association to traits being studied and for evidence of natural selection acting on these traits.

Genome-wide association studies (GWAS) allow inferring genotype-to-phenotype associations by detecting correlations between genetic variation and trait differences based on linkage disequilibrium (LD), i.e. the non-random association of alleles at different loci in a given genomic region (Bush and Moore, 2012). These genome scans can have different conformations, depending on each case study. Divergence mapping, for example, relies on allele frequency data from two populations with different phenotypes, scanning these data to identify loci that are outliers of divergence relative to the background and can thus be locally adapted (Beaumont and Balding, 2004). This method was used, for example, to pinpoint loci involved in determining wing patterns in *Heliconius* butterflies (Nadeau *et al.*, 2012) or adaptation to freshwater in threespine sticklebacks (Hohenlohe *et al.*, 2010). However, if neutral divergence is high, this approach may be insufficient to differentiate outlier loci. In this case, admixture mapping can provide increased power, studying admixed individuals that express distinct phenotypes of interest. This method depends on the fact that, after some generations, LD between neutral and adaptive loci will decay to fine scales due to recombination (Crawford and Nielsen, 2013). This has been applied, for instance, to unravel regions underlying diseases in human populations (Price *et al.*, 2007) or male nuptial colour and body shape in sticklebacks (Malek *et al.*, 2012). Moreover, genome scans can be based on linkage mapping, based on experimentally controlled crosses between distinct parental populations. This approach is usually applied to map multi-locus traits through quantitative trait loci (QTL) mapping (Miles and Wayne, 2008); however, it only maps to large genomic regions due to few crossing generations, thus requiring additional testing (Miles and Wayne, 2008; Crawford and Nielsen, 2013). QTL mapping has been used, for example, to identify regions responsible for disease resistance in plants (Young, 1996) and colour patterns in *Heliconius erato* (Papa *et al.*, 2013).

Directional positive selection that favours genotypes responsible for advantageous traits is a mean for adaptation (Nielsen, 2005), which leaves particular signatures in the genetic variation at the affected loci. For example, if strong selection leads to the fixation of beneficial variants, their progressive increase in frequency leads to the local reduction of genetic variation in adjacent neutral regions linked to the locus under selection, a process known as selective sweep (Smith and Haigh, 1974; Kaplan *et al.*, 1989). As consequence, the levels of LD are expected to increase in regions under selection (Nielsen, 2005). This also results in changes in the distribution of alleles of existing mutations. Specifically, positive selection generally tends to increase the number of mutations segregating at higher frequencies when compared with a

neutral model (Nielsen, 2005). However, when selective sweeps occur, the distribution of alleles in the window of the genome affected by the sweep tends to be skewed for mutations segregating at either very low or very high frequencies (Braverman *et al.*, 1995; Fay and Wu, 2000), the latter being a unique pattern caused by the occurrence of sweeps (Fay and Wu, 2000). In addition, when positive selection acts only in some populations within a species, it can lead to increased differentiation among populations in the specific genomic regions subject to selection. At the population level, positive selection for adaptation can thus locally result in i) decreased genetic variation, ii) changes in the shape of frequency distribution of genetic variation, iii) extended LD, and iv) increased differentiation (Oleksyk *et al.*, 2010). Genome scans for genetic variation in populations can detect these deviations to neutral expectations, as seen, for example, for winter colour polymorphism in snowshoe hares (Jones *et al.*, 2018) or bill length in great tits (Bosse *et al.*, 2017).

Despite the successful application of these methods and the continuous drop of high throughput sequencing prices, NGS is still expensive (Metzker, 2010). Thus, studies relying on the use of many individuals can be hampered by the costs of sequencing the whole-genome at a high depth. Low-coverage sequencing is a popular alternative because it takes advantage of the trade-off between the number of samples and the sequencing coverage per individual, allowing variants to be accurately called through the combination of data obtained across samples (Le and Durbin, 2011; Li *et al.*, 2011). For instance, Pasaniuc *et al.* (2012) have shown that ultra-low-coverage sequencing (0.1 to 0.5X) captures almost as much common and rare variants as SNP arrays, achieving similar significance values in association statistics. Notwithstanding some limitations, such as the reliable inference of low frequency variants, this allows for an increased GWAS sample size, leading to increased statistical power while using a more cost-effective sequencing technique. Several studies have applied this strategy to pinpoint genotypes associated to multiple adaptive traits, including the flowering time of rice (Huang *et al.*, 2012), red colouration in canaries (Lopes *et al.*, 2016), or winter coat colour polymorphism in snowshoe hares (Jones *et al.*, 2018).

1.5. Museum collections as sources of genetic data

Natural History Museum (NHM) collections are a rich sample of biological variation at both inter and intraspecific levels, therefore constituting important resources to multiple fields of biodiversity research, including conservation, genetics, agriculture, and public health (Suarez and Tsutsui, 2004). These collections can be particularly relevant to the study of populations and species that are now extinct or highly endangered, or for which the collection of new

material is difficult or impossible (Payne and Soreson, 2002). Since the last century, researchers have been using NHM samples to study biodiversity, particularly, to evaluate the impacts on species' range and distribution caused by habitat loss, biological invasions, and climate change (Suarez and Tsutsui, 2004).

With the development of modern molecular techniques, specimens available at NHM became an important resource to conduct and improve genetic studies and have been used to answer questions related, for example, to phylogeny (Shapiro *et al.*, 2002), phylogeography (Larson *et al.*, 2005), changes in genetic variability over time (Bouzat *et al.*, 1998), and conservation (Poulakakis *et al.*, 2008). However, the low quantity and high fragmentation of endogenous DNA in these samples restrained the broad application of traditional population genetic methods (Wandeler *et al.*, 2007; Burrell *et al.*, 2015). More recently, the development of genomic techniques promoted the application of NGS to museum samples. Given that the first step of high throughput sequencing is to break the template DNA into short fragments, the size of DNA fragments from museum specimens, naturally fragmented, makes this first step unnecessary (Nachman, 2013; Burrell *et al.*, 2015). However, if DNA is too degraded, it might result in significant loss of data, due to extremely short library insert sizes (Burrell *et al.*, 2015). The application of NGS technologies with museum collections was first accomplished by Miller *et al.* (2009), who used hair samples from specimens of the extinct Tasmanian tiger to recover the species' mitochondrial genome. Later on, Rowe *et al.* (2011) used skin and skull samples from *Rattus norvegicus* specimens to perform a whole-genome sequencing approach. Nonetheless, both works had methodological difficulties, particularly, a moderate degree of contamination and low mapping quality, respectively. The reliable collection of high-quality genomic data from NHM specimens was only later achieved (e.g. Bi *et al.*, 2013; Staats *et al.*, 2013), together with the first application of these data to conduct population-level analyses (Bi *et al.*, 2013).

Generally, although some NGS methods are not applicable to museum samples (e.g. RNA-sequencing), several high throughput approaches are feasible, particularly whole-genome sequencing and targeted sequence capture re-sequencing (Burrell *et al.*, 2015), and some studies have been conducted to tackle technical questions related to this application (e.g. Bailey *et al.*, 2016; Lim and Braun, 2016; McCormack *et al.*, 2016). This led to the development of a new field in genomics, known as museum genomics. Some recent works from this field explored, for example, changes in the genomic diversity of alpine chipmunks (Bi *et al.*, 2013), the radiation of guenons (Guschanski *et al.*, 2013), the phylogeography of Caribbean rodents (Fabre *et al.*, 2014), or the systematics of a grass lineage endemic to Madagascar (Silva *et al.*, 2017). Given that NHM collections can house numerous specimens representative of local adaptation, they hold great promise to study the genetic basis of adaptive traits.

1.6. Objectives

Understanding the genetic mechanisms that underlie the evolution of key adaptive traits is a central question in evolutionary biology. Hence, exploring the underpinnings of seasonal coat colour change and winter coat colour polymorphism in all seasonal coat colour changing species is important to dissect the evolution of these traits across species. Additionally, in a context of climate change, polymorphism in winter coat colour may be an important source for evolutionary rescue and, therefore, understanding the genetic mechanisms underlying winter coat colour variation can help monitor the evolution of the trait.

In this work, museum samples collected at the Swedish Museum of Natural History, that cover the polymorphic zone of *Mustela nivalis* in Sweden, were used in an NGS framework that aimed to genetically characterise this population and to identify the genomic regions involved in the determination of winter coat colour variation. Specifically, this work aimed at:

1. Identifying the genetic basis of the winter coat colour polymorphism in the least weasel, pinpointing candidate genes to underlie distinct winter colour phenotypes;
2. Inferring genomic signatures of natural selection that elucidate the adaptive value of variation in winter coat colour;
3. Clarifying the structure of the population of least weasels from Sweden, providing preliminary data on its history and association with winter coat colour distribution.

This study is expected to provide strong insights into the genetic basis of seasonal coat colour variation and to highlight the importance of natural history museum collections as valuable resources to characterise the architecture of local adaptations.

2. Materials and Methods

2.1. Sampling

In this work, 179 samples of *Mustela nivalis* were analysed, which were partitioned to address distinct objectives. These samples included i) 83 samples from specimens originally collected in Sweden and obtained from the collection of the Swedish Museum of Natural History (Stockholm, Sweden), ii) 63 samples from Poland, obtained from the collection of the Mammal Research Institute, Polish Academy of Sciences (Białowieża, Poland), and iii) 33 samples obtained from the Zoological Research Museum Alexander Koenig (Bonn, Germany), which resulted from the crossing experiments of Frank (1985). All samples consisted of patches of dried skin, with original collection dates ranging between December of 1954 and April of 2010.

Samples were classified as *nivalis* or *vulgaris* morphs according to the line dividing the dorsal and ventral pelage in the summer colouration (straight – type I, *nivalis*; ragged – type II, *vulgaris*), or as *nivalis* if presenting a winter-white colour. Samples from Sweden corresponded only to winter specimens (with collection date registered between November and March), 41 with brown coat and 42 with white coat. Some of these brown coat specimens (N = 13) presented a colouration type I, for a total of 55 *nivalis* and 28 *vulgaris* individuals. In the literature, the change to white in winter and colouration type I (i.e. two traits associated with the *nivalis* morph) have been described as being co-inherited, suggesting a single genetic basis or tightly linked genetic determinants (Frank, 1985). The discordance between the type of demarcation line and winter coat colour could be due to dissociation of the traits, plasticity in the timing of the moults or errors in the indication of the collection date in the museum database. Hereafter, *nivalis* will designate individuals with straight line or winter-white colour, and *vulgaris* will refer to individuals with ragged line and winter-brown colour. The samples from Poland included 15 winter-white samples and 49 summer samples with both types I and II phenotypes. The geographic distribution of samples from Sweden and Poland is shown in Figure 5. Samples from the breeding experiments correspond to the three types of crosses performed by Frank (1985). Specifically, tissues were retrieved from i) 5 F1 *vulgaris* specimens resulting from the crossing of the parental individuals (one *nivalis* female and one *vulgaris* male), ii) 20 individuals (including *nivalis* and *vulgaris* specimens) resulting from the backcross of F1 *vulgaris* males with the parental female, and iii) 8 *nivalis* specimens resulting from the crossing of the parental female with *nivalis* males born from the backcross. Detailed information about all samples can be found in Tables S1 – S4 (Appendix I).

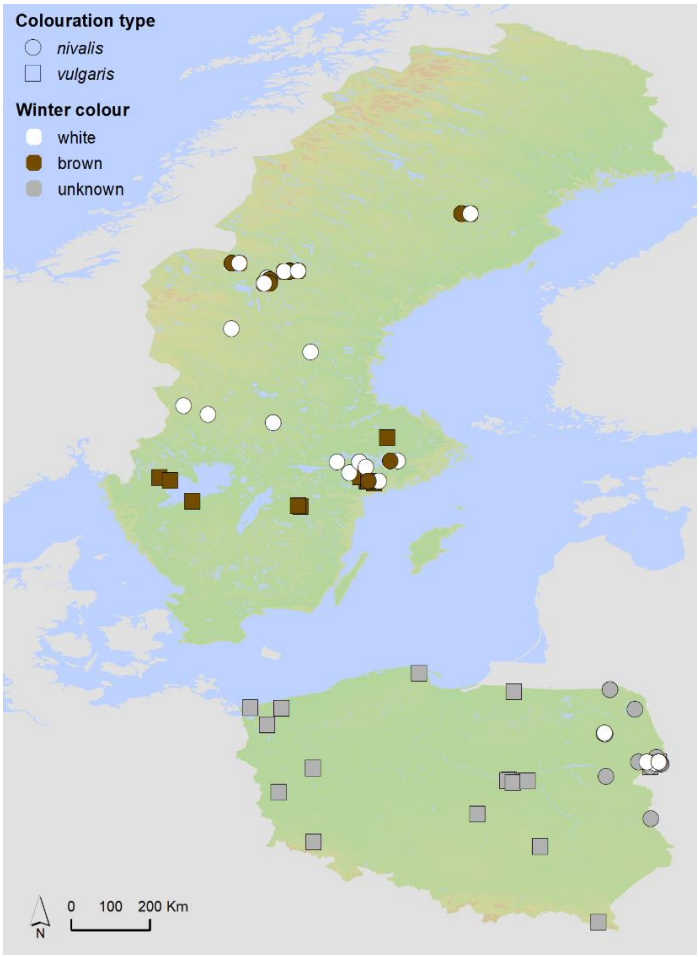

Figure 5 – Distribution of *Mustela nivalis* samples from Sweden and Poland included in the study. Each point represents one sampling location. Information on the samples collected at each location is given in Appendix I. The map was created using ArcMap v.10.1 (ESRI).

2.2. DNA extraction

DNA extractions for all samples used in this work were conducted in an isolated and autonomous low-DNA-status room, appropriate for extractions of low-quality/quantity DNA samples, with exclusive and sterilised equipment to prevent contamination. Dry tissues were hydrated in deionised water during one to two days prior to extraction, which was conducted following the protocol of Dabney *et al.* (2013). Multiple extraction rounds were performed, with a maximum of 15 samples per round, and a negative control was used in each of them to test for contamination. After each round, DNA concentration was measured with a Qubit Fluorometer (Invitrogen). For each sample, DNA concentration was measured twice, and the mean value was considered for further steps.

2.3. Library preparation and sequencing

Individually barcoded DNA libraries were initially prepared for 59 Swedish samples, following the protocol of Meyer and Kircher (2010) with the modifications described in Kircher *et al.* (2012). Briefly, an initial step of blunt-end repair was performed to eliminate overhanging DNA ends. The USER® enzyme (New England BioLabs, Inc.) was included in the blunt-end repair, in order to remove uracil bases resulting from cytosine deamination, which can occur in historical DNA sequences (Briggs *et al.*, 2010). Standard Illumina adapters were ligated to both ends of the molecule and nicks were removed through a fill-in reaction. Next, an indexing polymerase chain reaction (PCR) was performed to add sample-specific indexes. During this step, the following additional modifications to the protocol were applied: i) the AccuPrime™ Pfx DNA polymerase was used due to its proofreading activity, and ii) six replicates were performed to increase the complexity of the genomic representation. Additionally, each library was double-indexed to reduce inaccuracies in sample identification (Kircher *et al.*, 2012) and to allow pooling all samples in a single sequencing run. Final libraries were quantified by qPCR and were diluted to a concentration of 18 pM, as required for sequencing.

Sequencing of the 59 samples was conducted in a MiSeq run (sequencing kit v3), producing 75 base-pairs (bp) paired-end reads, to estimate proportions of endogenous DNA, insert sizes per library, and PCR duplicate proportions. Estimates of the proportion of mapped reads were obtained with BamTools v.2.5.1 (Barnett *et al.*, 2011) and used as a proxy for the proportion of endogenous DNA. The distribution of insert sizes was estimated through custom-made scripts and PCR duplicates were estimated using the MarkDuplicates function from PicardTools v.1.141 (software available at http://broadinstitute.github.io/picard/). These estimates were used to predict sequencing efficiency for subsequent higher throughput sequencing runs. Forty samples were selected for subsequent sequencing, 20 of each winter coat colour, following several criteria. First, to assure adequate assignment of the colour morphs, only specimens with collection date registered between December and February, at the height of winter coat colour (King, 1979), and with at least 95% of the body with brown or white colour (following Mills *et al.*, 2013) were selected. Second, specimens captured along and near the contact zone described by Stolt (1979) (as cited in King and Powell, 2007) were preferentially selected. Third, the number of males and females within and between morphs was balanced as possible, after fulfilling the previous criteria (Figure 6). Selected samples were next sequenced at low coverage in an HiSeq 1500 rapid run (sequencing kit v2), producing 100 bp single-end reads, which allowed re-calculating estimates of proportions of endogenous DNA, insert sizes per library, and PCR duplicate proportions, and adjust sample proportions pooled for the next sequencing run. The final sequencing run was performed in a full flow cell

(eight lanes) of HiSeq 1500 (sequencing kit v4), producing 100 bp single-end reads. Libraries were pooled in groups of five individuals, in a total of eight pools, and each pool was sequenced in one lane of the run. The number of individuals assigned to each winter colour was balanced in each pool, in order to prevent lane-specific throughput biases affecting only one of the colours. The combined sequencing effort aimed at approaching 20X coverage per colour morph.

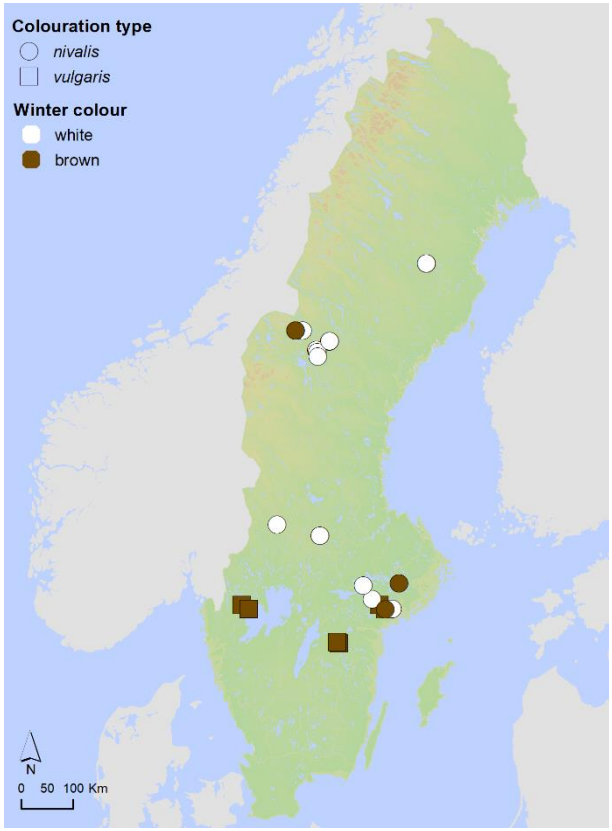

Figure 6 – Distribution of 40 *Mustela nivalis* samples used for whole-genome sequencing. Each point represents one sampling location. Information on the samples collected at each location is given in Appendix I. The map was created using ArcMap v.10.1 (ESRI).

2.4. Raw data trimming, mapping and processing

Raw sequencing data from the rapid run and eight lanes of HiSeq 1500 were cleaned by removing Illumina adapter sequences and low-quality reads with Trimmomatic v.0.36 (Bolger *et al.*, 2014). Reads were removed if phred-scaled quality scores were lower than 15 across windows of 5 bp and if read length was lower than 25 bp. The quality of the cleaned data was checked with FastQC v.0.11.7 (Andrews, 2015).

The ferret (*Mustela putorius furo*) reference genome (GenBank assembly accession no. GCA_000215625.1), the most closely related to *Mustela nivalis*, was used for mapping the sequenced reads. To improve mapping efficiency, an *M. nivalis* pseudoreference genome was first built. Briefly, all reads from the 40 individuals were first mapped to the ferret reference genome, and *nivalis*-specific fixed variants replaced in the reference. Using *pseudo-it* v.1.0.1 (Sarver *et al.*, 2017), three iterations were performed, and, for each of them, the reference obtained from the previous iteration was used as the new reference genome for alignment. This process is expected to minimise mapping bias that can arise due to the divergence between the sequenced reads and the reference and prevent lower mapping quality and loss of species-specific variation (Sarver *et al.*, 2017). Next, cleaned data were mapped to the pseudoreference using BWA-MEM, an algorithm implemented in BWA v.0.7.17 (Li and Durbin, 2009), with default parameters. Mapping statistics were calculated with BamTools v.2.5.1 (Barnett *et al.*, 2011).

Duplication estimates were obtained and duplicated reads were removed with the MarkDuplicates function from PicardTools v.1.141. This step ensures that each read used in the analyses derived from a distinct original DNA copy, which is particularly relevant because the library preparation procedure includes a PCR amplification step. Finally, the Genome Analysis Toolkit v.3.7.0 (McKenna *et al.*, 2010) was used to identify genomic regions of poor alignment with the RealignerTargetCreator function and to locally realign sequences in those regions with IndelRealigner. Final statistics per individual were calculated with Qualimap v.2.2.1 (Okonechnikov *et al.*, 2016).

2.5. Genotype likelihoods estimates and variant calls

To conduct population-level analyses, ANGSD v.0.921 (Korneliussen *et al.*, 2014) was used to estimate individual genotype likelihoods. This approach does not rely on genotype calls but takes into account the uncertainty associated with the genotype and incorporates it into the analysis. Because of the higher error rate of NGS data when compared with traditional sequencing techniques (Hawkins *et al.*, 2010) and of the higher difficulty in correctly inferring individual genotypes for sites covered by few reads (Harismendy *et al.*, 2009), the use of genotype likelihoods is particularly advantageous when working with low-coverage sequencing data (Kim *et al.*, 2011). In addition, to apply analyses that cannot incorporate genotype likelihoods, ANGSD was also used to infer genotypes at variable sites.

Genotype likelihoods were estimated with ANGSD, following SAMtools model (Li *et al.*, 2009) and filtering out reads with mapping quality lower than 20 and bases with phred-scaled

quality lower than 20. Reads with multiple mapping hits were also removed (-uniqueOnly). Major and minor alleles were inferred from genotype likelihoods and allelic frequencies were estimated following a maximum likelihood approach, filtering for a minimum minor allele frequency of 0.05. Estimates were conducted for sites covered in at least 75% of the individuals (-minInd 30) with a minimum of 1X coverage per sample. A maximum total sequencing coverage of 120X (i.e., approximately three times the mean global coverage) was established, in order to avoid overrepresentation of copy number variants or paralogous sequences that could result in the inference of false polymorphisms (Li, 2014; Schlötterer *et al.*, 2014). Polymorphic sites (SNPs) were inferred based on genotype likelihoods, using a polymorphism p-value threshold of 10^{-6} (Kim *et al.*, 2011). Additionally, SNPs were called based on genotype calls (-doGeno), estimating the posterior genotype probability with allele frequency as prior (-doPost 1) and retaining only sites with a posterior probability above 0.90 (-postCutoff).

2.6. Population structure and admixture

To infer the population structure of *Mustela nivalis* from Sweden, a principal components analysis (PCA) was performed with ANGSD, following a single read sampling approach. This method samples a single base from each individual at each position, thus being appropriate for low coverage data. SNPs called with ANGSD were first filtered to retain only one variable site at each 10 kb, to minimise non-independence of sites caused by linkage. Then, the consensus base per individual was sampled for each site (-doIBS 2), and the covariance matrix was inferred (-doCov 1). PC1 and PC2 were plotted with R v.3.3.2 (R Core Team, 2016). Next, population structure and admixture between populations were estimated with NGSadmix v.32 (Skotte *et al.*, 2013). Because this method incorporates genotype likelihoods directly into the analysis, ANGSD was used to produce the input file for NGSadmix, using genotype likelihoods. Sites were also filtered to keep only one SNP per 10 kb. Different K numbers of populations were tested, ranging from 2 to 10. For each K value, 200 independent runs were performed and the run with the highest likelihood was kept. Runs were optimised in a maximum of 2,000 iterations, with default convergence parameters, and only those achieving convergence were considered. Following Westbury *et al.* (2018), inferred admixture proportions for each K were considered meaningful if multiple independent runs converged at a likelihood score identical to the best run (i.e., with a likelihood difference lower than 0.1). The best K value was estimated using the 20 runs with the highest likelihoods of each tested K, following the method of Evanno *et al.* (2005) as implemented in Clumpak (available at http://clumpak.tau.ac.il) (Kopelman *et al.*, 2015).

To further understand the structure of this population, mitochondrial DNA (mtDNA) was also analysed. Individual mtDNA sequences were *de novo* assembled from cleaned sequencing reads. First, reads were mapped to *Mustela nivalis* mtDNA reference (NCBI accession no. NC_020639.1), using BWA with default parameters, to retain mitochondrial DNA reads (but possibly including true mtDNA and nuclear copies (NUMTs)). Then, ABySS v.1.9.0 (Simpson *et al.*, 2009) was used to assemble reads into contigs, generating different assemblies by varying k-mer size from 21 to 64 (maximum default k-mer size). Contigs longer than 1 kb from each assembly were aligned to *M. nivalis* reference, in order to confirm the consistency of the assembly across different runs. The consensus across runs was kept as the final sequence. Due to the high genetic variability of the control region of mtDNA (Moritz *et al.*, 1987), which can result in poor assembly, this region was removed from final sequences. Only positions represented in all specimens were kept for the analyses. Estimates of mitochondrial genetic diversity were calculated with DnaSP v.6.11.01 (Rozas *et al.*, 2017). Then, a median-joining haplotype network was constructed using Network v.5.0.0.3 (Bandelt *et al.*, 1999) to assess the relationship between distinct haplotypes. To assess the significance of mtDNA structure, individuals were partitioned according to three distinct criteria: i) the genetic populations inferred from PCA and NGSadmix analyses, ii) the winter coat colour (winter-white or winter-brown), and iii) the colouration type (*nivalis* or *vulgaris*). Exact tests of differentiation were conducted with Arlequin v.3.5.2.2 (Excoffier and Lischer, 2010), considering a significance threshold of $p < 0.05$.

2.7. Population phylogeny and demography

To further explore the history of the *M. nivalis* populations from Sweden, individuals were attributed to one of three populations, according to the results from both NGSadmix and PCA analyses. For samples with inconsistent assignment between analyses, NGSadmix results were followed, given that they incorporate all genetic variability, whereas PCA results are only based on the two major principal components. Specimens with mixed contribution from different population units, as inferred with NGSadmix (< 0.85 assignment probability), were not considered in this analysis.

Treemix v.1.13 (Pickrell and Pritchard, 2012) was used to infer the pattern of population splits. First, allele frequencies for each population were calculated with ANGSD. Read-level filters used were the same described for the genotype likelihood estimations. Allele frequencies were estimated following a maximum likelihood approach (-doMaf 1), using the original ferret reference genome to determine the ancestral state. Calculations were performed for sites

represented in i) all individuals for the two smaller populations (N = 5 and N = 7) and ii) at least half of the individuals for the larger population (N ≥ 9), with a minimum of 1X coverage per sample. Maximum coverage per site was established at three times the mean coverage of each population (17X, 22X and 51X, from the smallest to the largest). Next, a set of custom-made scripts was used to prepare the input for Treemix. Briefly, for each population, ancestral and derived allele counts were calculated based on the previously estimated allele frequencies. The ferret reference genome was added to the dataset as a fourth population, being used as outgroup. Data of the four populations were combined, keeping only sites represented in all populations. Additionally, invariable sites among the three *M. nivalis* populations and the ferret were removed, thus keeping only SNPs or substitutions relative to the original ferret genome. Positions where the minor allele count was lower than 2 in all *M. nivalis* populations were also removed, to ensure a minimum minor allele frequency of at least 0.05 across all populations. Finally, sites were filtered to retain one at each 20 kb, to avoid non-independence due to linkage. In Treemix, the best tree topology was inferred following a maximum likelihood approach. Sample size correction was turned-off (-noss option) to avoid overcorrection due to the small size of three populations, and a global round of rearrangements was done after adding all four populations to the tree (-global). To confirm the consistency of the results, 50 independent runs were performed, and the tree topology represented in at least 70% of the runs was kept as the maximum likelihood tree.

Even though our population sequencing data with low individual coverage are not adequate to properly estimate the complete site frequency spectrum (SFS), due to difficulties in inferring low frequency variants and the extent of missing data (Pool *et al.*, 2010; Buerkle and Gompert, 2013), the demographic history of the least weasels from Sweden was tentatively explored. Coalescent simulations based on a composite likelihood approach were conducted with fastsimcoal2 v.2.6.0.3 (Excoffier *et al.*, 2013), to infer demographic parameters based on the joint site frequency spectrum (jSFS). ANGSD was used to estimate the jSFS for each population pair. First, site allele frequency (SAF) likelihoods per population were estimated based on individual genotype likelihoods (-doSAF 1), using the original ferret reference to determine ancestral states. Read and site-level filters were the same applied when estimating allele frequencies for determining the population phylogeny. To estimate the jSFS, resulting SAF files were optimized in 50 independent runs per population pair, following a maximum likelihood approach on the realSFS program, part of the ANGSD package, and the run with the highest likelihood was kept as the best optimisation. Next, different demographic models were fitted to the jSFS, using fastsimcoal2. The population tree inferred with Treemix was used as underlying topology, and eight models were fitted, varying the levels of migration: i) no migration events, ii) migration between one pair of populations, iii) migration between two pairs

of populations, and iv) migration between all population pairs (see Figure S1 – Appendix II). Migration events were assumed asymmetric and population sizes kept constant through time. Model parameters were estimated assuming a general mammalian mutation rate of 2.2×10^{-9} per base per year (Kumar and Subramanian, 2002). For each model, 50 independent runs were performed, with 100,000 simulations per likelihood estimations (-n) and with 50 cycles of the likelihood maximisation algorithm (-L). The best model was chosen by a comparison procedure based on the Akaike Information Criterion (AIC) (Akaike, 1974). Comparisons of the empirical jSFS with the expected under the best demographic model were used to evaluate the fit of the model to the data, using *∂a∂i* v.1.7.0 (Gutenkunst *et al.*, 2009).

2.8. Whole-genome scans for association

To identify candidate regions of the genome to determine winter colour phenotypes, scans of differentiation along the genome were performed, under two distinct approaches, to identify regions with extraordinary allele frequency changes between morphs. Given the incomplete overlap between the winter colour and type of demarcation line in the sample set, genome scans were conducted twice, following two different criteria, i.e. contrasting i) winter colours (winter-brown vs. winter-white), and ii) colouration types (*nivalis* vs. *vulgaris*). This resulted in groups of 20 winter-white, 20 winter-brown, 24 *nivalis*, and 16 *vulgaris* samples.

First, a population-based approach was conducted in Popoolation2 v.1.2.0.1 (Kofler *et al.*, 2011). Initially, individual bam files were merged with SAMtools v.1.6 (Li *et al.*, 2009) to the group assignment of each specimen. Resulting files were filtered for reads with a minimum mapping quality of 20 and used to generate one mpileup file per comparison (winter colour or colouration morph), also with SAMtools. Only sites represented in a minimum of six individuals per group were considered, to decrease variances associated with extremely low sample sizes. Next, this file was modified and filtered for a minimum base quality of 20 with the *mpileup2sync.pl* script of PoPoolation2. Indels were identified in the mpileup with the *identify-indel-region.pl* script, defining a window around indels of 5 bp, and the output file was used to remove indels from the main file with the *filter-sync-by-gtf.pl* script. Differentiation between groups was measured through computation of the fixation index (F_{ST}), following the approach of Karlsson *et al.* (2007) as implemented in PoPoolation2 (--karlsson-fst). F_{ST} was calculated in sliding windows (*fst-sliding.pl*), considering only sites with a minimum total minor allele count of 3 and coverage between 10 to 60X per population. Multiple values were run for both window and step sizes, testing both overlapping and non-overlapping windows. Similar results were obtained with all combinations tested, and results are presented in non-overlapping 10 kb

windows. Results were filtered for windows with a minimum of 20 SNPs and plotted with R v.3.3.2 (R Core Team, 2016), using the qqman package (Turner, 2018). The threshold for selecting potential regions of interest was established at the 0.01% windows with higher F_{ST} values, and candidate regions were defined where at least two consecutive windows were identified above the established threshold. Additionally, Fisher's exact test (Fisher, 1922) was used to test the significance of differences in allele frequencies. Estimations were conducted in sliding windows, using the *fisher-test.pl* script and applying the same minor allele count and coverage filters as for F_{ST}. Multiple window and step-sizes were also tested, and results are presented using the same window size and SNPs filter mentioned above. Significant values were established following a Bonferroni-corrected p-value < 0.05.

Second, a case-control test of association was performed with ANGSD, using a likelihood ratio test of allele frequency differences between either winter-brown (case) and winter-white (control) or *vulgaris* (case) and *nivalis* (control) individuals. Allele frequencies were estimated based on individual genotype likelihoods, using the read-level filtering options described for genotype likelihood estimations. SNPs were inferred based on genotype likelihoods and filtered for sites represented in at least six individuals per phenotype. The obtained likelihood ratio statistic conforms to a χ^2 distribution with one degree of freedom and was used to compute p-values. Results per SNP were plotted with R, and significant associations were established following a Bonferroni-corrected p-value < 0.05.

Regions of high differentiation identified in at least one of the described approaches were further investigated. Using Ensembl (https://www.ensembl.org/) annotations for the reference genome and gene predictions from the NCBI database (https://www.ncbi.nlm.nih.gov/), the genetic content of the selected regions was inspected to identify candidate genes to underlie different winter phenotypes. Particular attention was given to candidate genomic regions including genes with functions related to coat colour determination, circadian rhythms, or regulation of these and other related processes important for seasonal moulting, as described in the literature (e.g. Bennett and Lamoreux, 2003; Slominski *et al.*, 2004; Hoekstra, 2006; Hubbard *et al.*, 2010).

2.9. Genotyping of SNPs and functional impact

A more thorough analysis of the existent SNPs was performed along the genomic region with the strongest allele frequency differences between winter phenotypes. Allele frequencies were calculated for both winter-white and winter-brown individuals with ANGSD, based on genotype likelihoods and using the ferret original reference to determine the ancestral allele.

Considering the results from Frank's (1985) experiments (see Introduction), the winter-white phenotype is expected to be recessive and thus winter-white specimens to show fixation of the causal variant. Winter-brown specimens are expected to be either heterozygote or homozygote for the alternative allele, which results in polymorphism in a population perspective. Therefore, SNPs in this candidate region and flanking 100 kb were filtered according to the respective allele frequencies, selecting only those which were: i) fixed for one allele in winter-white individuals, and ii) fixed or polymorphic for the alternative allele in winter-brown individuals (frequency ranging between 0.4 and 1). SNPs that met these criteria were further validated by visual inspection with Integrative Genomics Viewer v.2.4.1 (Thorvaldsdóttir *et al.*, 2013). Furthermore, their genomic location was examined to determine if it was i) a genic or intergenic region and ii) if genic, a coding or non-coding region.

From the resulting catalogue, SNPs were selected for genotyping, favouring variants within protein-coding genes and a regular representation along the candidate region. In total, 9 SNPs and one associated 12 bp deletion were selected (Figure 7), resulting in a total of 8 fragments to amplify. Genotyping of all selected variants was conducted for the total dataset from Sweden. Two variants were further genotyped in the specimens from Poland and from the crossing experiments of Frank (1985).

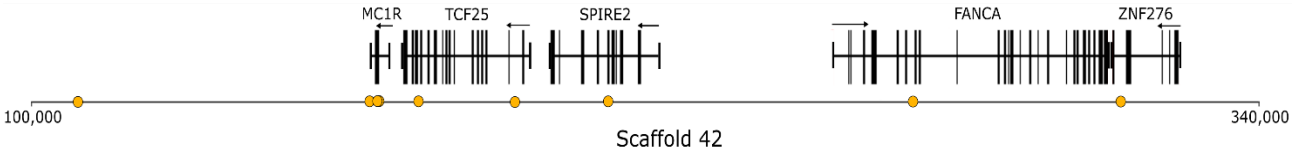

Figure 7 – Location of the variants selected for genotyping along the candidate region of scaffold 42 (GL896939.1). Each yellow dot represents a SNP. The structure of candidate genes identified in this region is also shown.

Primers to amplify the selected variants were designed based on the pseudoreference and each pair was tested for melting temperature and primer-dimer using NetPrimer (PREMIER Biosoft; available at https://www.premierbiosoft.com/netprimer/). Because of the expected DNA degradation of museum samples, primers were designed so that the amplicons would not exceed a length of 150 bp. PCR reactions were optimised using temperature gradients and testing for distinct primers and DNA concentrations. The presence of unspecific amplicons was checked through electrophoresis, and final PCR conditions were selected based on the best PCR band without unspecific amplicons. All pre-amplification steps were carried out in an isolated PCR room, reserved for handling low-quality DNA samples. PCRs were conducted in a T100-Cycler (Bio-Rad) and a negative control was used in every reaction to test for

contamination. PCR products were analysed through an electrophoresis and only samples with a visible band in the gel were used for sequencing. PCR products were cleaned following an ExoSAP protocol, to remove unincorporated primers and nucleotides that might compromise sequencing (Bell, 2008). Next, a cycle sequencing reaction was performed, and products were cleaned to remove unincorporated nucleotides. Finally, cleaned products were dried and rehydrated with formamide. Further details on primers' sequences (Table S5) and conditions for each step are available in Appendix III. Sanger sequencing was conducted in a 3130XL Genetic Analyzer (Applied Biosystems), resorting to capillary electrophoresis.

Resulting sequencing data were visualised with FinchTV v.1.4.0 (Geospiza, Inc.; Seattle, WA, USA; http://www.geospiza.com) and DNA sequence chromatogram data were visually corrected when needed. Particularly, sequences from heterozygote individuals for the deletion were manually edited to obtain both alleles. For each amplicon, sequences were imported to BioEdit v.7.2.5 (Hall, 1999) and aligned to the pseudoreference, using Clustal W (Thompson *et al.*, 1994) as implemented in BioEdit. The genotype of each specimen was verified, and the significance of the association between phenotypes and genotypes was tested through Fisher's exact test (Fisher, 1922).

The functional impact of an amino acid substitution caused by candidate SNPs in *MC1R* gene was inferred using SIFT (Ng and Henikoff, 2003), as implemented in the web server http://sift-dna.org (Sim *et al.*, 2012). SIFT predictions are based on the assumption that if an amino acid is conserved across orthologous protein sequences its change is more likely to have an impact on protein function. The Sequence tool was used to search the substitutions against both UniProt-SwissProt and UniRef90 databases, using a median sequence conservation value of 3.00 and removing sequences more than 95% identical to the sequence of interest. Sequences from both databases were considered for comparison only if annotated as "Melanocyte-stimulating hormone receptor" or "Melanocortin receptor 1". Functional changes were assumed if the normalised probability of tolerated change was lower than 0.05.

2.10. Tests of selection

To test for evidence of positive selection in candidate regions identified from the whole-genome scans in the population from Sweden, a selective sweep analysis was conducted with SweeD v.3.3.4 (Pavlidis *et al.*, 2013), a software that detects sweeps based on site frequency spectrum (SFS) patterns of SNPs. The analysis was performed independently for the four groups defined for the whole-genome scans (winter-white, winter-brown, *nivalis*, and *vulgaris*). First, allele frequencies were estimated with ANGSD, based on genotype likelihoods and using

the same read-level filters described for estimations of genotype likelihoods. SNPs were polarised relative to the original ferret reference genome and retained only if represented in at least six individuals. The input file was prepared based on these estimations, using a custom-made script. SweeD was run for the scaffolds of interest, including monomorphic sites and estimating the composite likelihood ratio (CLR) at each 10 kb along the scaffold. Regions in the top 1% higher CLR values were identified as putatively selected. Because the existence of structure among specimens of the same group could influence the detection of CLR outliers, the analysis was repeated with the three populations inferred from the PCA and NGSadmix analyses, to verify the consistency of the results. The same parameters were applied, with exception to the smallest genetic population, where SNPs were required to be represented in five individuals, the total size of the population.

3. Results

3.1. Sequencing statistics

The sequencing output yielded a total of 2,014,968,153 raw reads, with an average number of raw reads per sample of 50,374,204. After quality trimming and removal of Illumina adapters, the average number of reads per sample was 50,124,206, with an average loss of 0.49% and resulting in a total of 2,004,968,248 reads.

After mapping the reads to the pseudoreference, initial estimates yielded an average of 86.67% of properly mapped reads per sample. This value corresponded to an average increase of 1.25% of mapped reads per sample (varying between a minimum of 0.90% and a maximum of 1.96%), when compared with mapping to the original ferret reference. Mean duplication values per sample were of 18.27% and, after removing duplicated reads, the percentage of properly mapped reads dropped to 84.28%. After all filtering steps, mean individual sequencing coverage was of 0.98X (see sequencing statistics in Table 1 and detailed estimates per sample in Table S6 – Appendix IV). Total mean genome coverage was of 39.03X, corresponding to 18.60X in winter-white specimens and 20.43X in winter-brown individuals, or to 22.78X in *nivalis* and 16.25X in *vulgaris* specimens.

Table 1 – Summary of sequencing statistics. Mean values for each parameter are calculated per individual. Minimum and maximum values are also presented.

		Mean	Minimum	Maximum
Number of reads	Raw data	50,374,204	43,211,612	62,478,645
	Trimmed data	50,124,206	43,019,389	61,666,982
Mapping (%)	Before removing duplicates	86.67	79.91	90.23
	After removing duplicates	84.28	76.82	88.68
Duplication (%)		18.27	15.65	22.81
Coverage (X)		0.98	0.65	1.33

3.2. Population structure, admixture and phylogenetic history

To understand the genetic relationships between specimens from the population of least weasels from Sweden, a PCA was initially conducted, based on 100,132 putatively independent SNPs (at least 10 kb apart), and the results suggest the existence of three genetic

populations. The first principal component (PC1; 3.72% of explained genetic variance) separates one larger group (A) containing mostly winter-white/*nivalis* individuals from two other groups (B and C) of mostly winter-brown/*vulgaris* specimens, whereas the second principal component (PC2; 3.62% of explained genetic variance) splits these latter populations. Exceptions to this distribution include i) nine winter-brown specimens, of which six are *vulgaris* type, included in group A and ii) one *nivalis* individual in group B. Moreover, some individuals from group A are placed in a more intermediate position, suggesting the possible occurrence of admixture between these three populations (Figure 8).

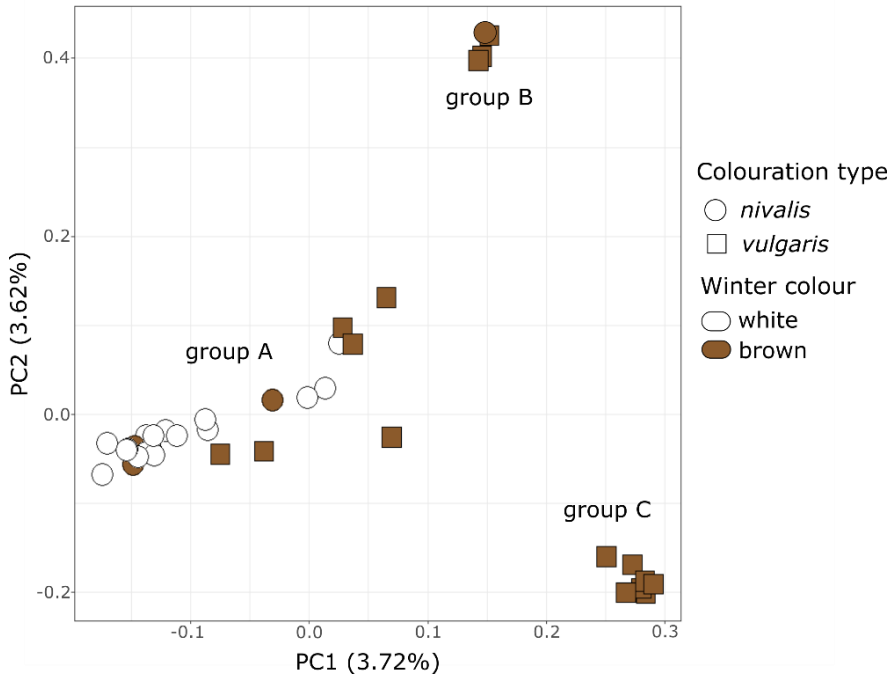

Figure 8 – PCA plots based on SNP data for *Mustela nivalis* from Sweden. The first (PC1) and second (PC2) principal components are shown.

Analyses of ancestry proportions were conducted with NGSadmix, based on 172,642 putatively independent SNPs. Convergence of the estimates across independent runs for each K was only achieved for K values of 2 and 3 (Figure 9A). For K = 2, one of the clusters comprises individuals from the northernmost part of the distribution and the northern part of the polymorphic zone (PZ) of winter colours (Figure 9A and C), whereas the other corresponds to the southern and part of the eastern region of the PZ. For K = 3, the subdivision of this second cluster according to the southern and eastern location of the samples is supported. Results for higher values of K are shown in Figure S2 (Appendix V). K = 3 was estimated as the best partition (Figure S3 – Appendix V), in accordance with the PCA results (Figure 9B).

For convenience, these three clusters were named according to their geographic distribution: Northern (N = 18), Eastern (N = 5), and Southern (N = 7), corresponding, respectively, to groups A, B, and C of the PCA. This analysis suggests the occurrence of individuals of putatively admixed ancestry (assignment proportion < 0.85) in the area where both winter colours or colouration types exist (Figure 9C).

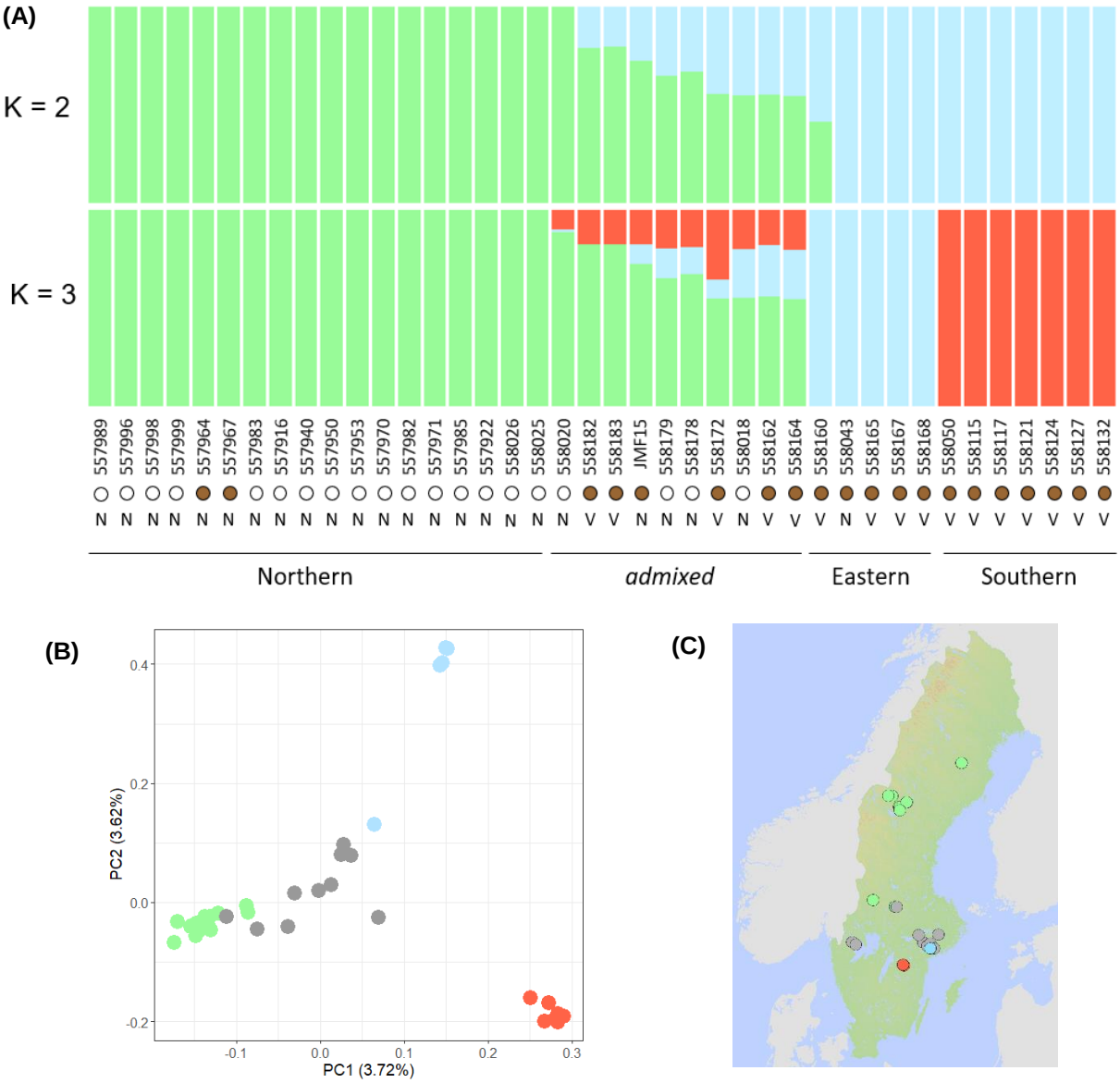

Figure 9 – Admixture proportions of *Mustela nivalis* from Sweden. (A) Ancestry proportions for K = 2 and K = 3, inferred with NGSadmix. The winter colour (brown or white) and colouration type (*nivalis* – N or *vulgaris* – V) of each specimen are also depicted. (B) PCA results for SNP data, colouring samples according to NGSadmix results for K = 3. (C) Geographical distribution of each lineage inferred for K = 3. Inferred putatively admixed individuals (with assignment probability < 0.85) are depicted in grey in (B) and (C).

Complete mitochondrial DNA sequences were recovered for 38 individuals (the sequence of specimen 558132 was only partially recovered, and no sequence was obtained for specimen 558127). Network analysis was performed using an alignment of 11,549 bp, which included 39 individuals. A total of 143 variable sites and 28 haplotypes were identified, resulting in levels of haplotype diversity (Hd) of 0.977 ± 0.013 and nucleotide diversity (π) of 0.00221 ± 0.00009. Results suggest the existence of six major haplogroups, from which four are mainly composed of individuals from the Northern population, one is mainly composed of Southern and Eastern specimens, and another is composed of putatively admixed specimens. One Eastern specimen is included in one of the Northern haplogroups, while the remaining admixed specimens are spread among three Northern and the Eastern/Southern groups (Figure 10A). Considering the winter colour or colouration type partitions, one haplogroup is mostly composed of winter-brown/*vulgaris* specimens, two are present only in winter-white/*nivalis* individuals, and two are mostly winter-white/*nivalis*. The remaining haplogroup is composed of two *nivalis* specimens, each with a different winter colour (Figure 10B-C).

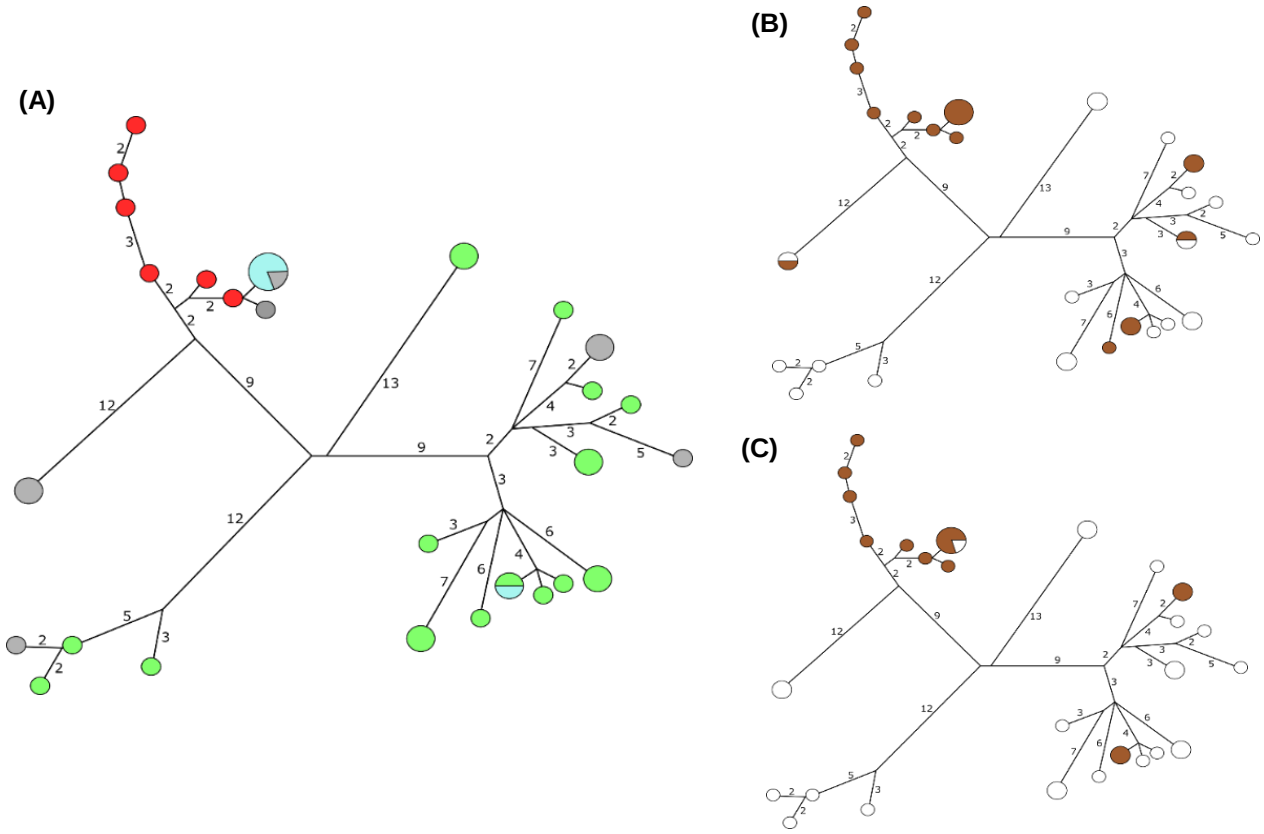

Figure 10 – Mitochondrial DNA haplotype network of *Mustela nivalis* from Sweden. Each circle represents one haplotype, with size proportional to the number of individuals sharing that haplotype. Branch length is proportional to the genetic distance between haplotypes, and numbers in each branch indicate the number of mutated sites. Colour codes correspond to the (A) genetic population (Northern – green, Southern – red, Eastern – blue, or putatively admixed – grey), as inferred with NGSadmix, (B) winter coat colour (white or brown), and (C) colouration type (*nivalis* – white or *vulgaris* – brown) of each individual.

The significance of mtDNA structure was estimated with Arlequin. Differentiation results suggest that there is no significant structure when analysing samples partitioned according to the identified genetic populations ($p > 0.05$). The same result was obtained with or without the inclusion of admixed individuals. Contrastingly, when mtDNA was analysed considering winter colours or colouration types, results indicate significant haplotypic structure ($p < 0.05$).

To gain deeper insights into the population history of *Mustela nivalis* in Sweden, the phylogeny of the three genetic populations was inferred using Treemix, based on a total of 70,089 putatively independent sites. The resulting maximum likelihood tree supports a first split of the Northern population (composed of *nivalis* specimens, mostly winter-white), while the Southern and Eastern populations (composed of winter-brown specimens, mostly *vulgaris*) group together (Figure 11).

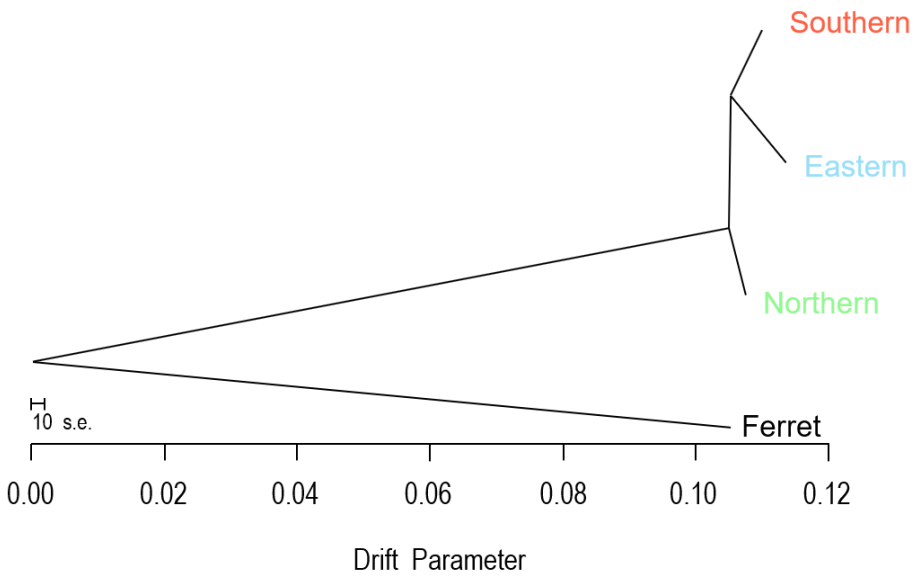

Figure 11 – Maximum likelihood phylogeny of *Mustela nivalis* from Sweden inferred with Treemix. The ferret reference genome was used as outgroup. The scale bar shows 10 times the standard error (s.e.) of entries in the covariance matrix of samples.

Additionally, using fastsimcoal2, we attempted to model population history based on the joint site frequency spectrum (jSFS), which could give important information about parameters of population divergence, such as effective population sizes, times of divergence, or rates of gene flow (e.g. Hernandez et al., 2007; Gutenkunst et al., 2009; Nielsen et al., 2009). However, the empirical jSFS optimised from the data showed poor fit to allow robust inferences, which was confirmed when plotting and comparing jSFS modelled under the best demographic scenario (see Tables S7-S8 – Appendix VI for details on the selection and parameters of the best model) to the empirical optimisation. The wide distribution of model residuals (Figure 12) suggests that the modelling could not properly capture the empirical variation, resulting in unreliable inferences (Gutenkunst *et al.*, 2009). For this reason, we opted not to consider these analyses.

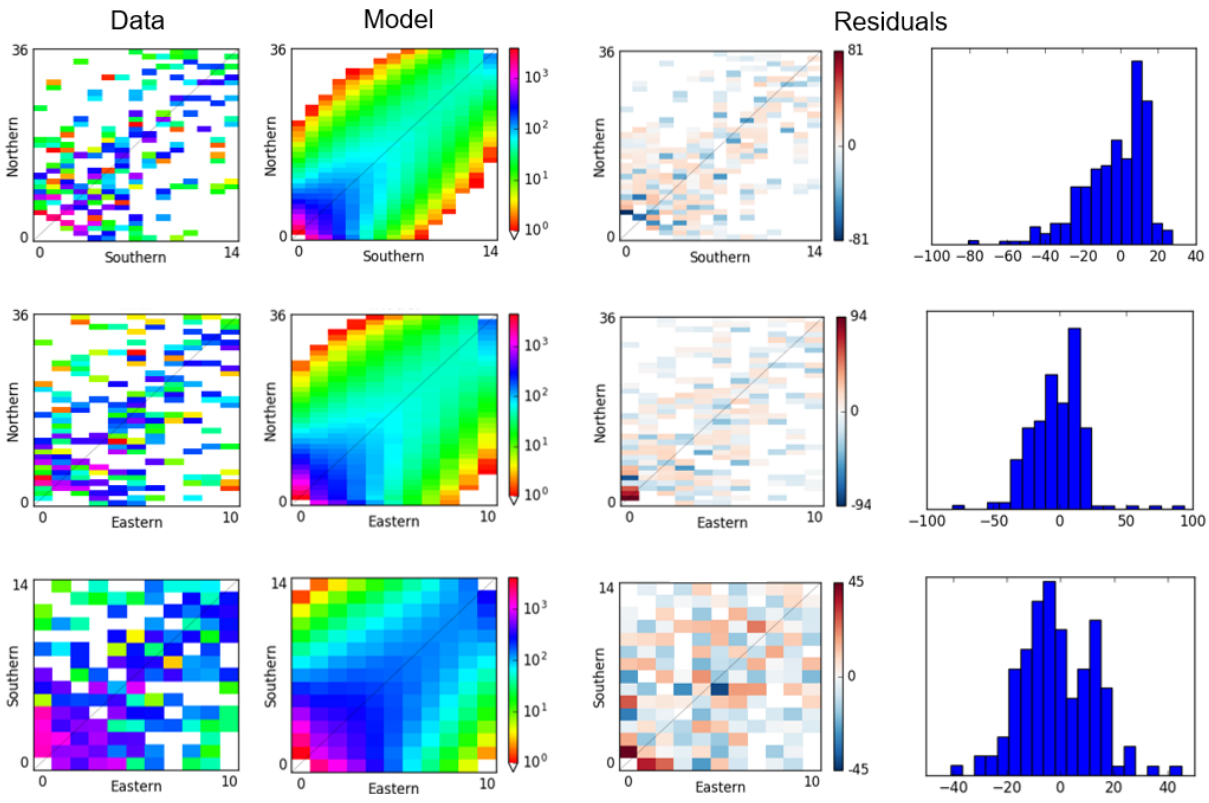

Figure 12 – Fit of the best demographic scenario to the observed data, for each pairwise population comparison. From left to right, each column represents i) the observed jSFS for each pair of populations as estimated from the empirical data, ii) the expected jSFS under the best demographic scenario for each pair of populations, iii) model residuals for each jSFS, and iv) histograms with the distribution of model residuals.

3.3. Whole-genome scans for association

To uncover the genetic underpinnings of the winter phenotypic differences between individuals of *Mustela nivalis*, whole-genome scans of differentiation and association were conducted with Popoolation2 and ANGSD, considering both winter-white vs. winter-brown (hereafter WB) and *nivalis* vs. *vulgaris* (hereafter NV) contrasts. Globally, results were consistent between approaches.

In total, 14,444,119 and 14,069,754 SNPs were retained for F_{ST} estimates in WB and NV analyses, respectively. Mean genome-wide differentiation estimates (F_{ST}), calculated with putatively independent SNPs (10 kb apart), were of 0.0628 and 0.0651 for WB and NV scans, respectively. Highly differentiated windows were mapped mostly within scaffold GL896939.1 (scaffold 42 in Figures 13 and 14) in both WB (Figure 13A) and NV (Figure 14A) scans; however, windows with high differentiation identified in other scaffolds were inconsistent between approaches. The significance of the estimated differences in allele frequencies was tested through Fisher's exact test, in sliding windows. Outlier values were identified in both approaches (Figures 13B and 14B), with the region of scaffold 42 presenting the most extreme value for the NV comparison (Figure 14B), but none of them was identified as significant after correction for multiple tests. Also, case-control associations based on individual genotype likelihoods per SNP were tested with 1,257,810 SNPs for both WB and NV tests. No significant outliers were found in the WB test (Figure 13C) but one significant outlier ($p < 3.975 \times 10^{-8}$) was found in the NV test, in the same region of scaffold 42 previously identified as highly differentiated (Figure 14C).

Candidate genomic regions were defined if at least two consecutive windows were identified as highly differentiated (in the top 0.01% higher F_{ST} values) or if significant outliers were found in Fisher's exact test or case-control association tests. This resulted in three candidate genomic regions to be further explored at: i) scaffold 42 (GL896939.1: ~ 170,000 – 330,000), supported by both F_{ST} estimates and the case-control association test in the NV grouping, ii) scaffold 206 (GL897103.1: ~ 2,220,000 – 2,240,000), supported by F_{ST} estimates in the NV contrast, and iii) scaffold 234 (GL897131.1: ~ 1,390,000 – 1,420,000), supported by F_{ST} estimates in both NV and WB comparisons.

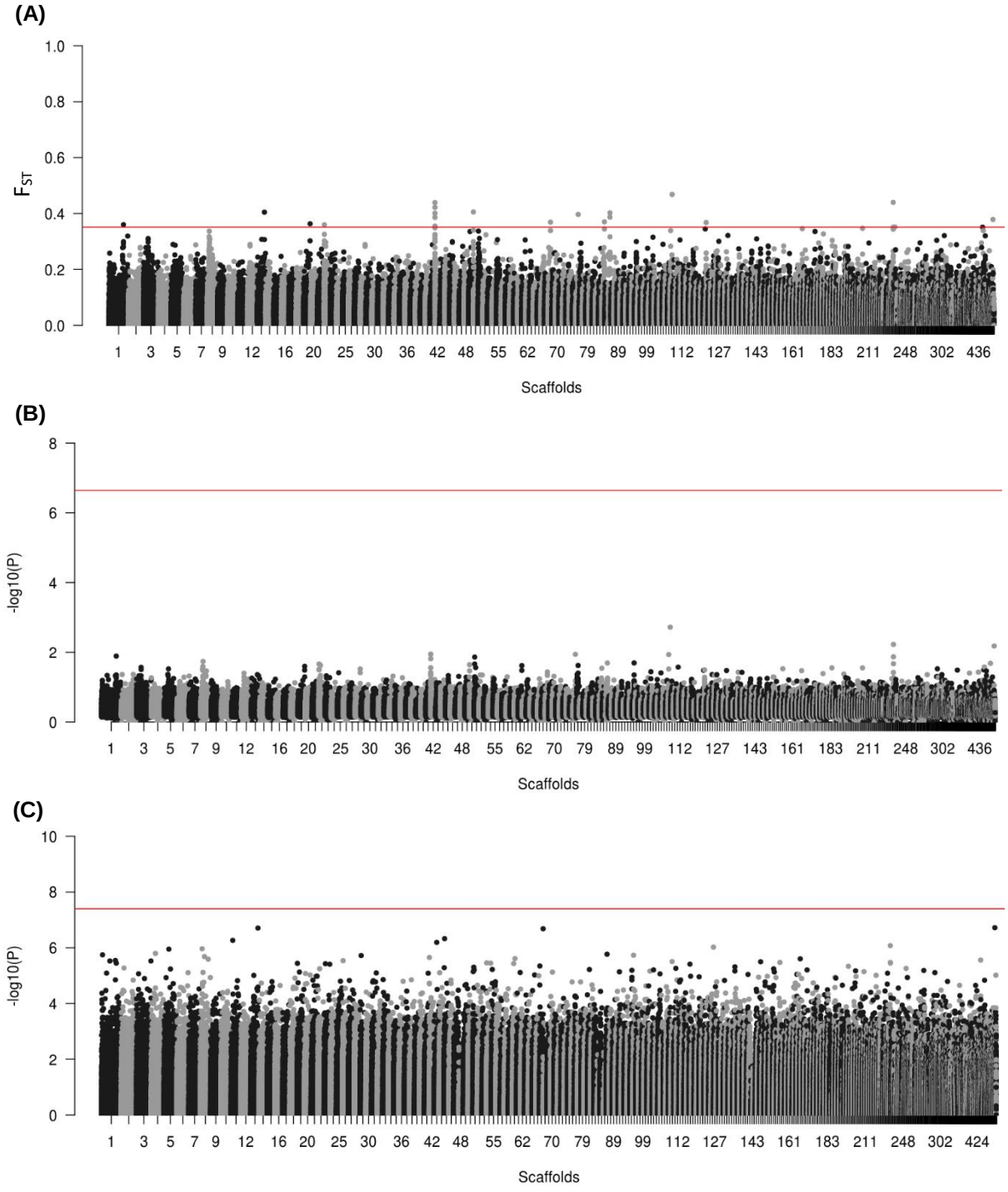

Figure 13 – Whole-genome scans conducted between winter-white and winter-brown least weasels. (A) F_{ST} values between winter colour morphs, averaged for 10 kb non-overlapping windows. The red line marks the 99.99th percentile of windows with highest F_{ST} values. (B) Significance of allele differences between populations as estimated through Fisher's exact test, averaged for 10 kb non-overlapping windows. No significant outliers exist following a Bonferroni-corrected $p < 0.05$. (C) Likelihood ratio test for differences in allele frequency between winter-white and winter-brown specimens calculated for individual SNPs. No significant outliers exist following a Bonferroni-corrected $p < 0.05$.

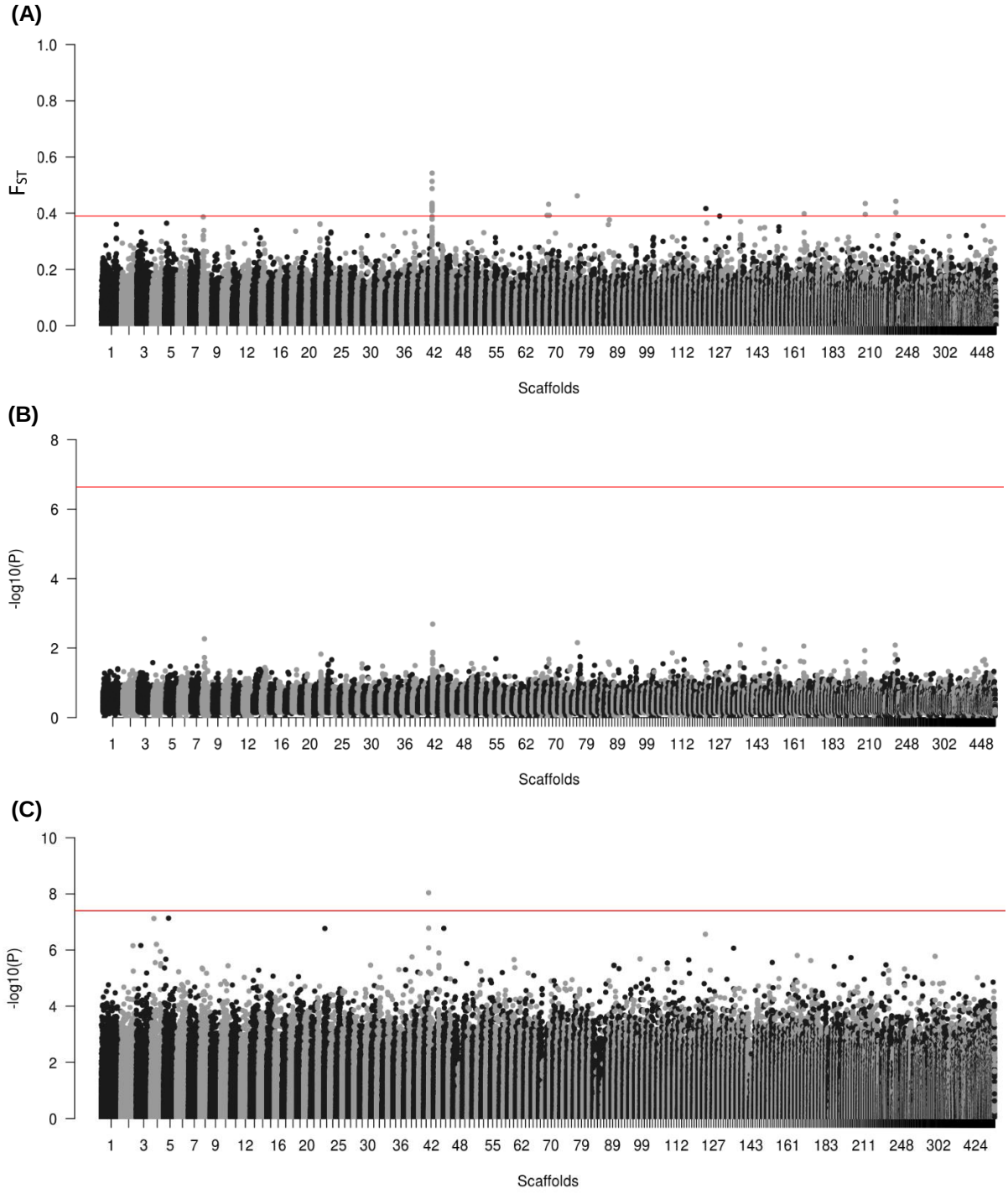

Figure 14 – Whole-genome scans conducted between *nivalis* and *vulgaris* least weasels. (A) F_{ST} values between colouration types, averaged for 10 kb non-overlapping windows. The red line marks the 99.99th percentile of windows with highest F_{ST} values. (B) Significance of allele differences between populations as estimated through Fisher's exact test, averaged for 10 kb non-overlapping windows. No significant outliers exist following a Bonferroni-corrected $p < 0.05$. (C) Likelihood ratio test for differences in allele frequency between *nivalis* and *vulgaris* specimens calculated for individual SNPs. SNPs above the red line are significant outliers following a Bonferroni-corrected $p < 0.05$.

The genetic content of windows with elevated differentiation/association identified in at least one of the approaches was further investigated. At the candidate region in scaffold 42, five genes were identified: i) *MC1R*, a well-known pigmentation gene (Hoekstra, 2006), ii) *TCF25*, important for embryonic development (Olsson *et al.*, 2002), iii) *SPIRE2*, involved in intracellular vesicle transport along actin fibres (Schuh, 2011), iv) *FANCA*, involved in DNA repair (D'Andrea and Grompe, 2003), and v) *ZNF276*, that may be implicated in transcriptional regulation (Laity *et al.*, 2001). No genes were identified at scaffold 206, whereas, at scaffold 234, *DRG2* gene was identified, that has an important function in cell growth (Ko *et al.*, 2004) (Table 2, but see Table S9 – Appendix VII for the complete list of genes identified in other highly differentiated windows).

Table 2 – List of genes identified within the three candidate regions defined from genome scans. For each gene, information is given about the encoded protein, the scaffold where it was identified, and the location of the complete gene sequence within the scaffold. Positions are given according to the ferret reference genome.

Gene	Encoded protein	Scaffold	Position
MC1R	Melanocortin-1 receptor	GL896939.1	173,236 - 177,472
TCF25	Transcription factor 25	GL896939.1	179,741 - 203,544
SPIRE2	Spire-type actin nucleation factor 2	GL896939.1	207,322 - 227,671
FANCA	Fanconi anemia complementation group A	GL896939.1	260,028 - 311,891
ZNF276	Zinc finger protein 276	GL896939.1	312,270 - 325,401
DRG2	Developmentally regulated GTP-binding protein 2	GL897131.1	1,398,432 - 1,418,511

3.4. SNPs genotyping along a candidate region

From the candidate genomic regions identified, one appeared particularly relevant in the context of this work because it was consistently identified across whole-genome scans and it contains a well-known pigmentation gene – *MC1R* – and another with described functions related with colouration – *SPIRE2*. Selected SNPs were genotyped to obtain high-quality individual genotypes and examine the segregation of these variants in relation to the trait of interest. In total, 82 SNPs were identified conforming to the expected allele frequencies (Table S10 – Appendix VIII) and, from those, 9 SNPs and one associated deletion were selected for genotyping to verify the association with both morphs (Table 3).

Table 3 – Variants selected for genotyping experiments. For each variant, position relative to the ferret reference genome, type of variant, ancestral and derived alleles, and genomic location are indicated. Alleles are shown in the 5' – 3' direction of the ferret reference genome, and the ferret variant was defined as ancestral.

Position	Type	Alleles		Genomic Location	PCR amplicon
		Ancestral	Derived		
118,608	SNP	G	A	Intergenic	W1
173,465	SNP	G	A	MC1R - 3' UTR	W2
175,008	SNP	A	T	MC1R - CDS	W3
175,009	SNP	G	T	MC1R - CDS	W3
182,432 – 443	deletion	GTGTACTTTTTA	–	TCF25 - Intron 16	W4
182,449	SNP	A	C	TCF25 - Intron 16	W4
200,524	SNP	A	C	TCF25 - Intron 3	W5
217,866	SNP	G	A	SPIRE2 - Intron 8	W6
275,086	SNP	G	A	FANCA - Intron 10	W7
313,924	SNP	T	G	ZNF276 - Intron 8	W8

UTR – untranslated region
CDS – protein coding sequence

For the specimens from Sweden, all 10 variable loci were genotyped. One *nivalis* specimen (code 558042) consistently failed PCR amplification and was thus removed from the dataset. For the remaining 82 individuals, amplification and sequencing were generally successful, with three reactions (W2, W6 and W8) amplifying correctly for all individuals, whereas the others failed for one (W1, W3, W5), two (W7), or four (W4) specimens only. A significant association between phenotype and genotype was identified in all variants ($p < 0.05$, Fisher's exact test). For two adjacent SNPs identified in *MC1R* coding sequence (CDS) (positions 175,008 and 175,009), results show a perfect association between the colouration type and the expected inheritance pattern (Frank, 1985), i.e. *nivalis* specimens are homozygous, while *vulgaris* are either heterozygous or homozygous for the alternative allele. For variants increasingly distant from these SNPs, both upstream and downstream, a disruption of this pattern is verified, with multiple *nivalis* specimens presenting heterozygous genotypes (Figure 15).

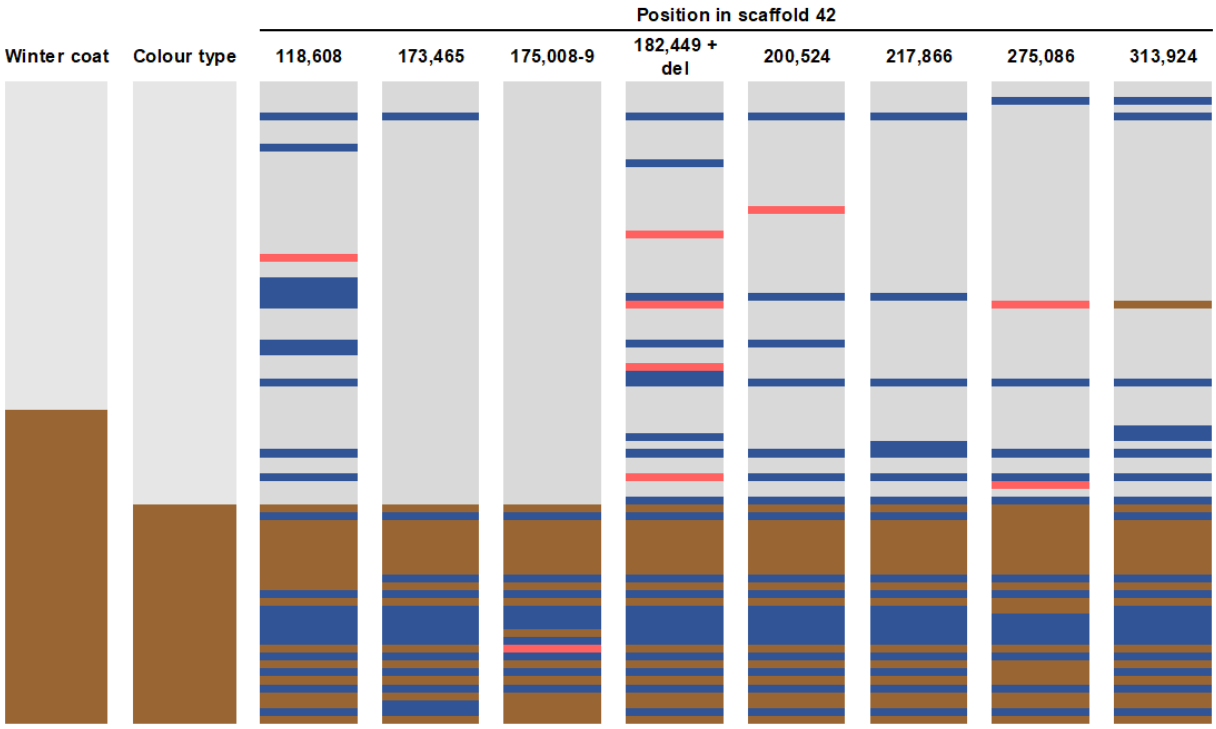

Figure 15 – Genotyping conducted in least weasels from Sweden. Each line represents one specimen, with its winter colour – white (in grey) or brown – and colouration type – *nivalis* (in grey) or *vulgaris* (in brown) – shown in the leftmost columns. For each variant, genotypes per specimen are indicated as follows: homozygous for allele predominant in winter-white/*nivalis* in grey, heterozygous in blue, and homozygous for the allele predominant in winter-brown/*vulgaris* in brown. Red lines indicate missing data.

To verify if the identified association remained in other least weasel populations, both SNPs in *MC1R* CDS were further genotyped in individuals sampled across a transition zone between colour morphs in Poland. From the 63 available specimens, genotyping was successful in all but one *nivalis* individual (code MRI51424). Again, results showed a perfect concordance between the colouration type of each individual and the expected pattern of inheritance, following the above-described association (Figure 16A).

Furthermore, both SNPs were genotyped in specimens from the interbreeding experiments of Frank (1985). For each individual, information was available regarding its colouration type, parents, and the type of crossing experiment from which it resulted (Table S4 – Appendix I). Therefore, following the obtained association pattern, the expected genotype of each specimen was assessed. Genotyping results were completely concordant with the expected inheritance patterns: i) the parental *nivalis* female and all backcrosses with *nivalis* type were homozygous for the allele predominant in *nivalis* individuals, and ii) all F1 individuals and backcrosses showing *vulgaris* type were heterozygous (Figure 16B).

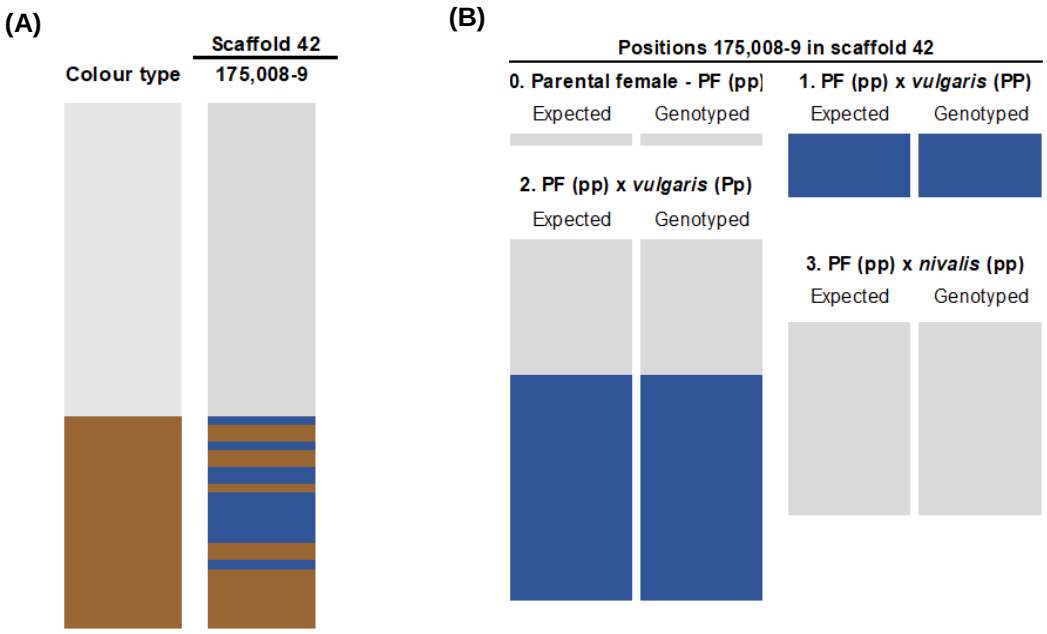

Figure 16 – Genotyping conducted in (A) least weasels from Poland and (B) individuals from the interbreeding experiments of Frank (1985). Each line represents one specimen and genotypes are indicated as in Figure 15. In (A), the colouration type (*nivalis* – grey; *vulgaris* – brown) of each individual is displayed in the left column. In (B), both the expected and obtained genotypes are shown per specimen. Individuals are grouped according to the type of crossing from which they result: 1. parental female (PF) with parental male; 2. F1 *vulgaris* males with PF; 3. backcross *nivalis* males with PF.

3.5. Functional impact of candidate variants

Given the complete association found in the SNPs from *MC1R* coding region, further inspection of the putative functional consequence of the difference was performed. Inspection of the location of these variants showed them to occur at the first two bases of a codon, resulting in one missense mutation. Consequently, whereas *nivalis* individuals always present a leucine residue at protein position 101, *vulgaris* alternative alleles result in the substitution of this amino acid by a lysine residue – L101K (Figure 17A). This substitution is located at the boundary between the second transmembrane domain (TMD2) and the first extracellular loop (EL1, Figure 17B).

Additionally, the functional impact of these variants was assessed with SIFT. The altered protein sequence was compared with 79 and 49 orthologous sequences from the UniProt-SwissProt and the UniRef90 databases, respectively. Results consistently indicated the missense mutation as potentially non-tolerated (normalised probability of tolerated change of 0.00 and 0.02, respectively), thus suggesting a possible change of protein function.

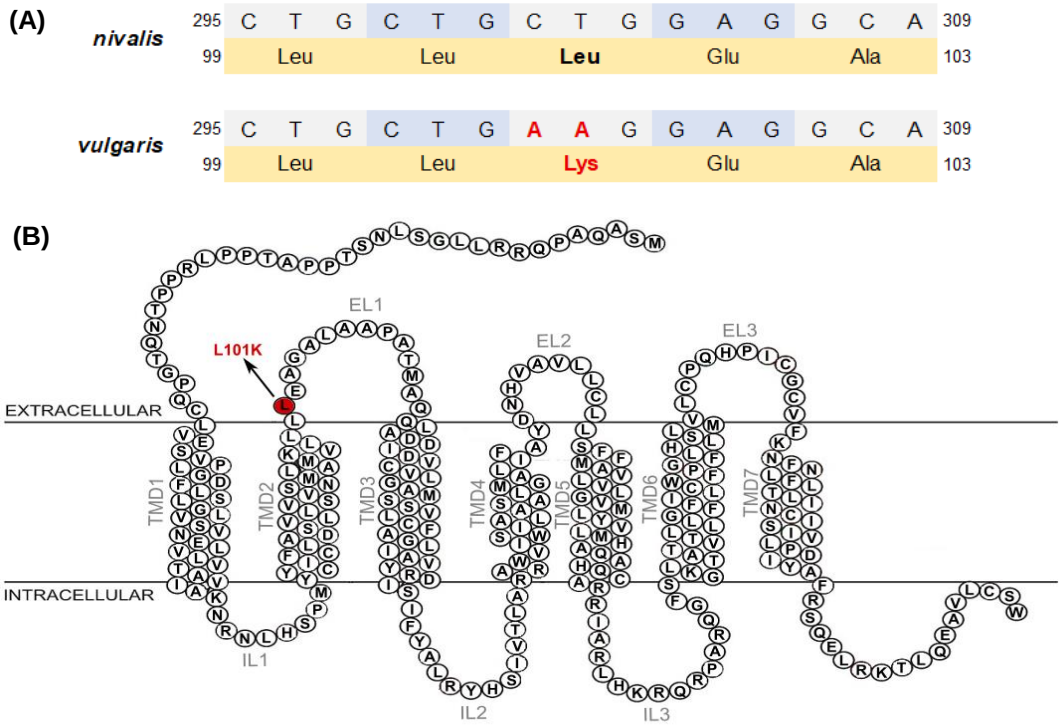

Figure 17 – Candidate variants to explain phenotypic differences in least weasels with distinct winter colours. (A) For each colour morph, both nucleotide and protein sequences in the region of *MC1R* SNPs are displayed. Nucleotide and amino acid positions are indicated according to the coding region of ferret's *MC1R* reference sequence (XM_013056885.1, NCBI database) and to the ferret's protein reference sequence (XP_012912339.1, NCBI database), respectively. Variants identified in the *vulgaris* morph are highlighted in red. (B) Location of the identified missense mutation in the 2D-structure of MC1R. The *nivalis* protein sequence is shown. All transmembrane domains (TMD), extracellular loops (EL), and intracellular loops (IL) are indicated.

3.6. Tests of selection

Candidate regions identified in whole-genome scans in the populations from Sweden were additionally tested for positive selection, using SweeD, to understand if strong differentiation between phenotypes could be accompanied by recent selective sweeps, suggesting local adaptation. Tests were conducted for four groups (*nivalis*, *vulgaris*, winter-white, and winter-brown), and results are consistent between comparable groups (Figure 18). While no evidence of putative selective sweeps was detected along scaffold 42 for *nivalis* or winter-white individuals, results from the analysis detected the occurrence of an outlier region (in the top 1% CLR estimates) at the beginning of the scaffold, in both *vulgaris* and winter-brown specimens, that overlaps the identified candidate region. At scaffold 234, a putative selective sweep was identified ~ 40 kb after the candidate region, in *vulgaris* and winter-brown individuals. Another outlier at the same scaffold, ~ 110 kb before the candidate region, is also suggested for *nivalis* and winter-white specimens. No evidence of selective sweeps was found at scaffold 206 for any of the four groups. The same analysis was conducted for the three

genetic populations, to verify if grouping winter-brown/*vulgaris* individuals from different genetic populations was influencing the obtained signals. Results for the Northern and Southern populations were consistent with winter-white/*nivalis* and winter-brown/*vulgaris* individuals, respectively. However, no CLR outliers were detected for the Eastern population (Figure S4 – Appendix IX).

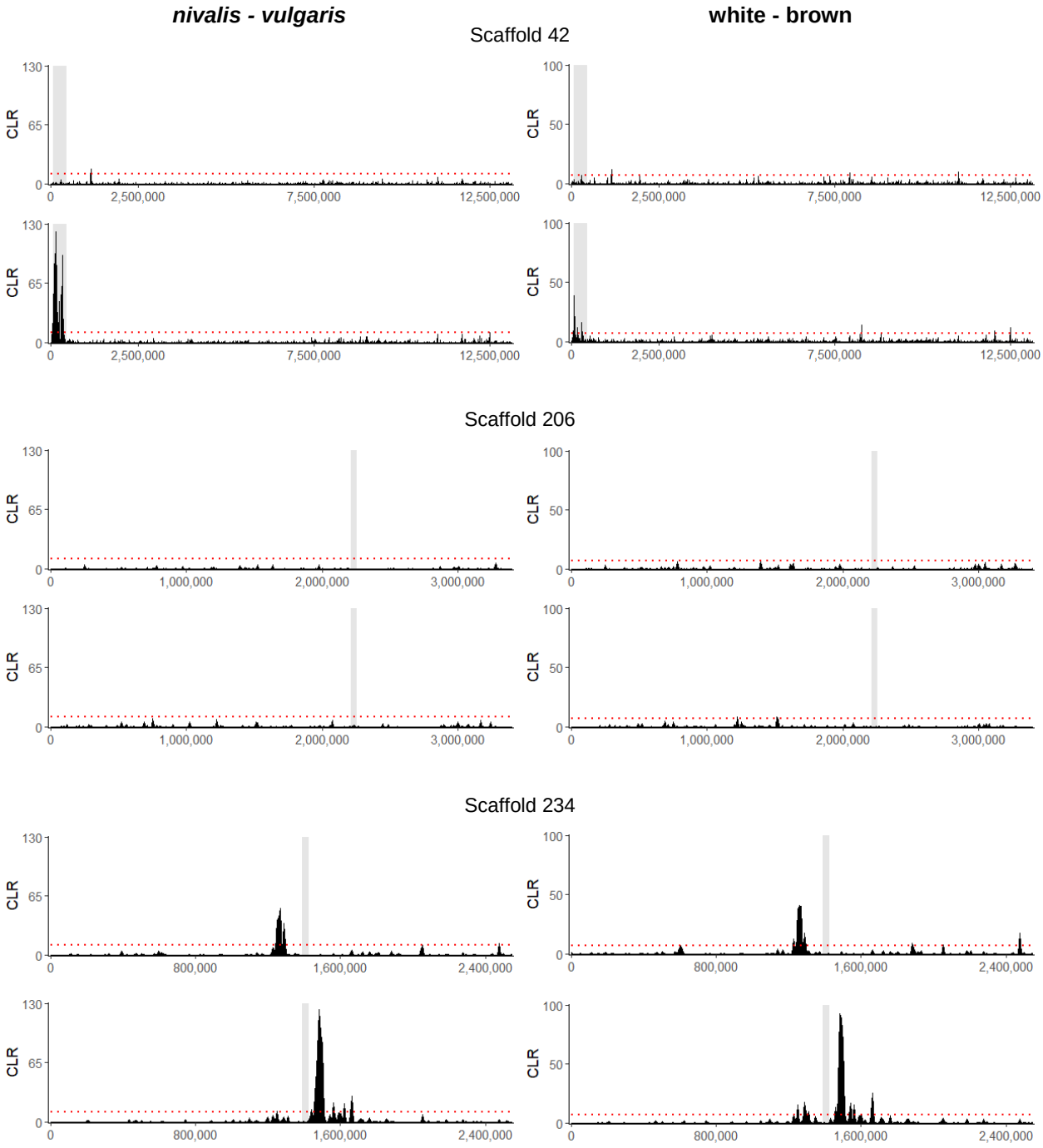

Figure 18 – Inferred selective sweeps across the three candidate regions at scaffold 42 (GL896939.1), scaffold 206 (GL897103.1), and scaffold 234 (GL897131.1). The left panel shows composite likelihood ratio (CLR) estimates for each scaffold in *nivalis* (top) and *vulgaris* (bottom) specimens. The right panel shows CLR estimates for each scaffold in winter-white (top) and winter-brown (bottom) specimens. Grey shades represent the identified candidate regions. Red lines represent the 99th percentile of estimated CLR values.

4. Discussion

Seasonal coat colour change is an important adaptation that allows boreal species to remain camouflaged in different annual environmental conditions and conceal from other animals, namely predators and preys. Recent research has suggested this trait is being severely challenged by snow cover reduction caused by global warming, leading to mismatches between the coat and background colours that disrupt crypsis and decrease species fitness (e.g. Zimova *et al.*, 2016; Atmeh *et al.*, 2018). Importantly, some of these species also have populations that maintain the brown colour year-round. These could act as a source of adaptive variants to winter-white populations, and polymorphic zones, where winter-white and brown morphs coexist, could potentiate evolutionary rescue (Mills *et al.*, 2018). Understanding the genetic basis underlying the evolution of winter coat colour variation is thus important, not only to understand the evolution of a relevant adaptive trait but also to follow its changes when responding to new selective pressures. Previous studies have pinpointed genes underlying distinct winter phenotypes in both the arctic fox (Våge *et al.*, 2005) and the snowshoe hare (Jones *et al.*, 2018); however, in most species with variation in winter colour, the mechanism underlying different phenotypes remains unknown.

In this work, we investigated the genetic basis of distinct winter colours in the least weasel, *Mustela nivalis*, using a population whole-genome sequencing approach applied to museum samples. We explored the genetic structure and history of the studied populations, in relation to winter colour variation. Next, we conducted scans to identify candidate genomic regions and genotyped SNPs along the strongest candidate region to obtain individual genotypes and verify their association with different morphs. Finally, we tested these regions for evidence of natural selection that could be indicative of local adaptation.

4.1. Population history of *Mustela nivalis* from Sweden

Understanding population structure underlying the genetic diversity in a model system is important because it provides a genomic context on which to interpret results obtained from the genome scans. For example, strong population structure would advise caution when interpreting results from scans of differentiation, given that extreme differentiation values could result not only from selection acting on genes responsible for phenotypic differentiation but also from population history (Excoffier *et al.*, 2009). Similarly, association mapping through case-control association studies can be confounded by population structure, leading to false

positive inference errors (Pritchard *et al.*, 2000). The PCA and NGSadmix analyses revealed the existence of at least three main genetic clusters in *Mustela nivalis* from Sweden, which are consistent with the geographical distribution of specimens (Figures 8 and 9). A northernmost population was identified, being composed mostly, but not exclusively, of winter-white/*nivalis* specimens, while the Southern and Eastern contained winter-brown/*vulgaris* individuals. These analyses also showed the existence of admixed individuals, thus suggesting the occurrence of gene flow between all three populations (Figure 9A). The evidence of admixture is consistent with a previous work that showed individuals harbouring distinct morphs to be completely interfertile (Frank, 1985). Gene flow is thus expected to reduce differentiation between groups of specimens with distinct winter colour phenotypes, and to reduce the confounding effects of population structure when identifying outlier loci that may be associated with the determination of the trait of interest. The structure of these populations was further analysed by exploring mitochondrial haplotypic structure. Higher levels of structure were detected within the Northern population when compared with whole-genome data, with the identification of four mitochondrial lineages, which may result from female philopatry (as in e.g. Purdue *et al.*, 2000). Contrarily, Southern and Eastern haplotypes are generally included in the same haplogroup (Figure 10A). Little haplotype sharing was verified between the latter populations and the Northern individuals, even though differentiation tests do not suggest significant structure, which can result from small population sizes that reduce the power to detect significant differences. Interestingly, some of the individuals inferred as admixed by the analysis of nuclear loci belong to a divergent haplotype, which may reflect additional population structure that was not detected by the most likely number of K populations from NGSadmix.

To better understand the history of these populations, demographic modelling was tentatively conducted; however, our data did not allow robust inferences of the joint site frequency spectrum (jSFS) to produce reliable inferences (Figure 12). Even though the simulations presented by Nielsen *et al.* (2012) suggest that jSFS could be optimised with accuracy from data with coverage as low as 1X, demographic modelling is particularly susceptible to missing data (Pool *et al.*, 2010). Several peculiarities of our dataset may have contributed to reduce the number of sites that could be used to infer the empirical jSFS reliably. These were, for example, the uneven coverage along the genome resulting from PCR-based libraries (e.g. Aird *et al.*, 2011; Dabney and Meyer, 2012; van Dijk *et al.*, 2014), or the need to apply stringent filters to have a proper representation of the populations with smaller sample sizes in the dataset. The ideal scenario to conduct a reliable demographic modelling would be based on large numbers of high coverage SNP data, with more precise inference of genotypes in the analysed individuals (Crawford and Lazzaro, 2012).

Overall, the analyses of population structure suggest that there is some but not total coincidence between the patterns of genetic differentiation and the distributions of both winter colours/colouration types. Additionally, one of the populations consistently includes individuals with both phenotypes, independently of analysing population structure considering winter coat colours or *nivalis*-*vulgaris* morphs: i) the Northern population includes both winter-white and winter-brown individuals and both *nivalis* and *vulgaris*, and ii) the Eastern population includes both *vulgaris* and *nivalis* specimens. These results suggest some disjunction between the two colour morphs and population history and are, to some extent, congruent with previous phylogeographic studies conducted within the species. For example, McDevitt *et al.* (2012) identified four mitochondrial lineages of *M. nivalis* occurring in Poland that explain current patterns of genetic variation; however, their distribution does not coincide with the distribution of *nivalis* and *vulgaris* individuals reported in the country (see Atmeh *et al.*, 2018). Likewise, Lebarbenchon *et al.* (2010) studied mitochondrial diversity in least weasels from the western-Palearctic region and found two major lineages, one comprising western specimens and the other, which can be divided into five subgroups, comprising eastern and Moroccan specimens. These patterns are, again, inconsistent with the European distribution of both *M. nivalis* morphs (see Abramov and Baryshnikov, 2000; King and Powell, 2007). Together, these results may suggest a role of local selective pressures in maintaining the geographic structure of the winter phenotypes, despite admixture between the morphs. However, our results still do not exclude that the origin of winter colours/colouration types results from ancestral population divergence and that admixture alone is responsible for the sharing of winter phenotypes in the same genetic population. Indeed, the population phylogeny (Figure 11) suggests that the Southern and Eastern populations (containing winter-brown/*vulgaris* individuals) are genetically closer. Also, in the PCA, the first principal component splits these specimens from the Northern population (containing mostly winter-white/*nivalis* specimens) (Figures 8 and 9B), whereas, in the admixture analysis, for K = 2, individuals are mostly divided according to their colouration type and winter colour (Figure 9A). Additionally, analyses of mtDNA structure suggest significant differentiation between individuals with distinct winter colours or *nivalis*-*vulgaris* morphs, with one of the identified haplogroups being predominant in winter-brown/*vulgaris* specimens and the remaining in winter-white/*nivalis* ones (Figure 10B-C). Therefore, further work will be needed to understand the origin of distinct winter phenotypes. For example, the inclusion of whole-genome data from a second population that covers a different transition zone between phenotypes and the analysis of the genetic variation of both populations would help clarify this question. Specifically, it could help understand if the maintenance of distinct winter colours results from historical divergence between two different morphs or local selective pressures irrespective of historical population structure (as in e.g. Jones *et al.*, 2018).

4.2. Insights into the genetic basis of winter colour polymorphism

4.2.1. Genome scans identify candidate regions for winter coat colour

To identify candidates to underlie differences in winter phenotypes, whole-genome scans of differentiation and association were conducted. The low-coverage sequencing approach here applied is expected to provide accurate estimates of intermediate allele frequencies (e.g. Parchman *et al.*, 2012; Buerkle and Gompert, 2013) but is known to have bias in estimating allele frequencies in the extreme of the spectrum (Buerkle and Gompert, 2013). Furthermore, it reduces the representation of individuals at each single SNP, thus reducing sample sizes, which we attempted to minimise by ensuring minimum representation of individuals in each phenotypic group. Still, the approach is expected to introduce artificial variance in the dataset, leading to background noise along the genome scans, which indeed we observed in our analyses (Figures 13 and 14). Some underlying population structure could be an additional confounding effect but whole-genome scans based on population allele frequency differences suggest low mean genome-wide differentiation estimates between winter colours/colouration types (F_{ST} = 0.06). These analyses identified clear outlier regions of differentiation, which were consistent between both F_{ST} and Fisher's exact test, which required that consecutive SNPs in the genome showed similar patterns of allele frequency differences, thus minimising the variance effect on individual SNPs. Still, the fact that no outlier window was found significant after correction for multiple tests in Fisher's exact test is probably due to this artificial variance. Additionally, the case-control test based on individual genotype likelihoods resulted in the identification of a single outlier SNP, which however coincided with one of the genomic regions identified in the analyses based on allele frequency differences. Despite the difficulties associated with working with degraded museum sampling and some possible confounding effects resulting from technical aspects, these analyses allowed the identification of three strong candidate regions of association, which mapped to scaffolds 42 (GL896939.1), 206 (GL897103.1), and 234 (GL897131.1). A particularly strong signal of differentiation was found in the candidate region from scaffold 42, in both winter white-brown (WB) and *nivalis-vulgaris* (NV) comparisons, and was consistently identified by different analyses.

The inspection of the genetic content of these regions allowed identifying candidate genes to cause differences in winter coat colour in the least weasel. Six genes were identified – *MC1R*, *TCF25*, *SPIRE2*, *FANCA*, and *ZNF276* at scaffold 42, and *DRG2* at scaffold 234 – from which two are known to be involved in processes that could explain differences in colouration patterns. From these two candidates, one is striking because of its widespread involvement in the determination of coat colouration phenotypes – *MC1R*, the melanocortin-1 receptor. The

interaction of this receptor with its two ligands – *agouti* signalling protein (ASIP) and α-MSH – determines the type of melanin that is produced (e.g. Hoekstra, 2006) and the colour of the pigment that is deposited in the keratinocytes of hairs, i.e. dark or light pigments (Slominski *et al.*, 2004; see Introduction). Mutations in this gene have been shown to determine numerous colouration phenotypes, both in natural and domestic populations (Hoekstra, 2006; Hubbard *et al.*, 2010). These mutations result in either i) increased activity of the receptor, resulting in eumelanisation of the pelage (e.g. Robbins *et al.*, 1993; Klungland *et al.*, 1995; Kijas *et al.*, 1998) or ii) decreased activity of the receptor, leading to the production of pheomelanin and consequent light coats (e.g. Kijas *et al.*, 1998; Everts *et al.*, 2000; Fontanesi *et al.*, 2006). Moreover, its involvement in the determination of differences in winter colour has also been reported (Våge *et al.*, 2005). *MC1R* thus appears as a strong candidate to underlie the winter colour polymorphism in *Mustela nivalis*. In addition, another of the genes in the candidate regions has known functions that can be related with colouration – *SPIRE2*. This gene encodes a modular protein with an important role in the nucleation of actin filaments (AFs) in organelle membranes (Dietrich *et al.*, 2013). Recent works have shown that this gene, together with *SPIRE1* and *FMN2* (Formin-2), is involved in an actin-dependent mechanism that allows long-range vesicle transport through AFs, for example, in mouse oocytes (Kerkhoff *et al.*, 2001; Schuh, 2011). Interestingly, it has also been shown that the interaction of myosin-Va (MYO5A) – a major vesicle transport motor – with AFs allows the peripheral dispersion of melanosomes inside melanocytes (Evans *et al.*, 2014). Together, these results could suggest the involvement of *SPIRE2* in the formation of AFs necessary for transport of melanosomes. In fact, Alzahofi *et al.* (2018) have recently reported that SPIRE1/2 and FMN1 (Formin-1) are fundamental proteins to generate the AFs needed for MYO5A dependent transport in melanocytes. Additionally, they propose that Rab27a – a protein that recruits MYO5A to melanosomes (Wu *et al.*, 2001) – can also recruit SPIRE proteins to the organelle membrane. Previous studies have shown that mutations on *MYO5A* and *Rab27a* genes are implicated in partial albinism or colour dilution in mice due to irregular pigment transfer from melanocytes to keratinocytes (e.g. Mercer *et al.*, 1991; Provance *et al.*, 1996; Wilson *et al.*, 2000). Therefore, because this gene is implicated in the same system, mutations in *SPIRE2* may also influence pigment deposition. If so, considering the close location of *MC1R* and *SPIRE2* within the candidate region at scaffold 42 (~ 30 kb apart), this gene could be also a candidate to underlie colouration differences between both morphs.

The relevance of the genes included in the additional outlier genomic regions of differentiation to the trait studied here must be further explored in the future, considering the function and ideally re-examining patterns of variation using a more contiguous genome assembly. The fragmentation of the used reference genome could, for example, divide signals

of extreme differentiation (Alkan *et al.*, 2010). For instance, mapping of the candidate regions to the dog (*Canis lupus familiaris*) reference genome (CanFam3.1 assembly) – the closest reference genome with known chromosomal structure – shows that both regions from scaffolds 42 and 234 map to chromosome 5, being ~ 22.4 Mb apart. While these can still be two independent signals given the distance and known chromosomal rearrangements between dogs and weasels (e.g. Ehrlich *et al.*, 1997), this shows how improving the quality of the reference genome would help improve the understanding of the inferred signals.

4.2.2. Candidate SNPs identified in *MC1R* coding sequence

The candidate genomic regions and genes to determine the colour traits in least weasels were identified using the extreme allele frequency differences inferred. However, the approach taken, based on low individual coverage, advises caution because the relationship between the segregation of the genotypes and the trait cannot be determined. This could only be accomplished inferring quality individual genotypes, using higher individual coverage (Nielsen *et al.*, 2011), for example, with targeted resequencing data from these regions (e.g. Otto *et al.*, 2010). This would allow confirming and narrowing down the identified associations. In this work, we have taken a step in that direction, by genotyping by sequencing SNPs identified along the strongest candidate region. This allowed i) understanding how genotypes associate with the trait of interest in this region, and ii) testing association in another transition zone and in a pedigree, therefore excluding possible confounding effects of local population structure, which could result in false positives.

From the loci genotyped in the population from Sweden, two consecutive SNPs in *MC1R* coding sequence showed perfect association with *nivalis* and *vulgaris* morphs, considering the dominance pattern proposed by Frank (1985) – dominant *vulgaris* and recessive *nivalis* (Figure 15). The exact same inheritance pattern was found in the transition zone in Poland and in the specimens from the crossing experiments (Figure 16), showing that the extreme differentiation found between colour phenotypes in Sweden is not a false positive resulting from local population structure. These results strongly suggest the universal involvement of *MC1R* in the determination of the distinct colour phenotypes in least weasels. However, it remains to understand whether the evolution of the trait coincides with historical divergence within the species or is independent of average population history. Understanding the population structure of the population of least weasels from Poland, in the context of the one from Sweden, will help clarify this question.

The perfect association of genotypes was found with the *nivalis-vulgaris* morphs. The 13 *nivalis* specimens with putative winter-brown coat that were genotyped in this work showed variation compatible with *nivalis* type specimens (i.e. homozygous for the predominant *nivalis* allele) and differed from the winter-brown *vulgaris* (heterozygous or homozygous for the predominant *vulgaris* allele). Summer colour types I and II (based e.g. on the dorsal-ventral demarcation line) and winter colour had been shown to be determined either by the same locus or by two tightly linked loci on the same chromosome (e.g. Frank, 1985; King and Powell, 2007). The observation of some *nivalis* specimens sampled in winter but showing brown coat seemed to favour the two-locus hypothesis, with putative admixture breaking the linkage between them. The genome scans were performed considering both the *nivalis-vulgaris* morphs and the classified winter coat colour (Figures 13 and 14) to account for possible differences in the genetic basis of the two phenotypes. To further explore these differences, F_{ST} estimates for WB and NV contrasts were superimposed and compared for the candidate region from scaffold 42. This was done to identify putative windows of increased differentiation specific to each comparison, which could support a distinct genetic basis of both traits (demarcation line and winter colour), within the major candidate region for association. F_{ST} estimates between distinct colouration types (i.e. the NV grouping) were consistently higher than the F_{ST} between distinct winter coats (Figure 19). This suggests that the differentiation is maximised when comparing *nivalis* and *vulgaris* types and is slightly reduced when the *nivalis* specimens classified as winter-brown are included in the winter-brown group.

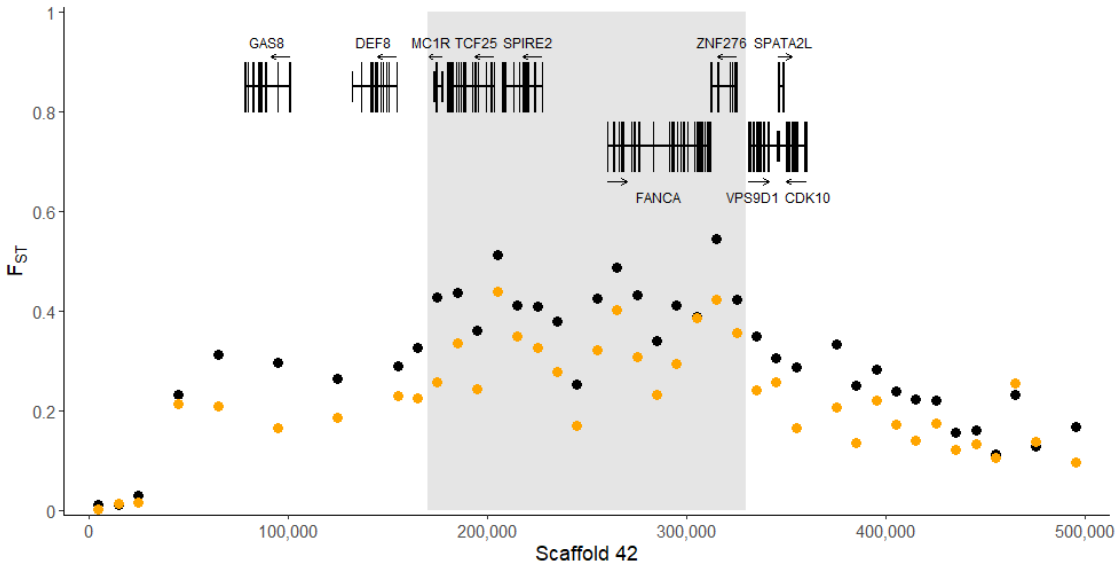

Figure 19 – Differentiation estimates across the candidate region from scaffold 42 (GL896939.1; grey shade), following the winter-brown vs. winter-white (orange dots) and *nivalis* vs. *vulgaris* (black dots) approaches. F_{ST} values are shown, averaged across 10 kb non-overlapping windows. The genic structure in this genomic interval is depicted above.

If two loci in the candidate scaffold are indeed responsible for demarcation line and winter colour, our low coverage sequencing approach and the low sample sizes of the discordant class of specimens may not have allowed sufficient power to identify the two loci, given their close physical proximity. Increasing the sample size of the apparent discordant cases could provide more resolution to separate the effect of the two loci in a differentiation scan. However, our current results do not support this separate effect. Also, we cannot exclude that the *nivalis* specimens we classified as winter-brown would actually change to white. First, plasticity in the initiation/termination of the moult is known to exist (e.g. Mills *et al.*, 2013; Zimova *et al.*, 2014; Atmeh *et al.*, 2018) and the sampled winter-brown *nivalis* could have not yet moulted to white or have already moulted back to brown. Even though specimens were selected at the height of the winter phenotype described by King (1979), Atmeh *et al.* (2018) have recorded specimens undergoing the spring moult in February, a month that was still included in our selection range. In fact, the collection dates of brown *nivalis* specimens are in early December or late January/mid-February, which are close to the limits of our selected seasonal range. Moreover, experiments conducted with a close relative, the stoat (*M. erminea*), have shown that warmer temperatures in autumn may extend the moult period from seventy hours up to 20 days, thus delaying in until December (Rothschild, 1942). Similarly, van Soest and van Bree (1969) have recorded the occurrence of stoats in complete brown pelage as late as December, for the autumn moult, and as soon as February, for the spring moult, in Netherlands. Delayed moult to white has also been sometimes observed in least weasels from Poland (Zub, personal communication), which would result in the inclusion of brown *nivalis* specimens in the inadequate winter colour group. This would explain the lower differentiation estimates in the WB grouping because these individuals, even in low number, would contribute to a slight homogenisation of allele frequencies of both groups. Second, the selection of specimens with each winter colour was done according to the information available in the museum database. Any eventual errors in registering the collection date of brown *nivalis* specimens (being actually from spring or summer), would cause the inclusion of these individuals in the wrong winter-colour group. Future research will address this question but, at this moment, our results do not exclude that a single locus may be responsible by both the demarcation line (type I or II) and winter colour, as previously suggested (e.g. Frank, 1985; King and Powell, 2007).

4.2.3. Possible functional impact of candidate causal variation

The involvement of *MC1R* in colouration differences is in-line with previous studies where the *MC1R-ASIP* system was implicated in the determination of distinct winter phenotypes. The first of those works was conducted in the arctic fox (Våge *et al.*, 2005). The authors identified

two cysteine substitutions occurring in *MC1R* coding region with a segregation pattern that was completely concordant with the dominant expression of the blue phenotype, which does not change to a winter-white colour, even though it also changes the expression of the summer phenotype. They hypothesised that these mutations would change the functioning of the receptor so that it would become constitutively activated, i.e. it would promote the production of darker pigment even in the absence of the agonist α-MSH (Lu *et al.*, 1998), thus being less responsive to changes of the *agouti*/α-MSH equilibrium that could trigger seasonal coat colour change. Jones *et al.* (2018) showed that variation in the seasonal expression of *ASIP* underlies different winter colours in the snowshoe hare. Genetic variation in a *cis*-regulatory region of the gene was implicated in decreased *agouti* expression in winter-brown individuals during the autumn moult. Therefore, the results from our work can reinforce former evidence of the role of *MC1R-ASIP* interaction in explaining intraspecific variation in winter colour.

Inspection of the candidate SNPs in *MC1R* coding region showed them to cause a non-synonymous mutation, that leads to the substitution of a leucine residue by a lysine at position 101 (L101K). Additionally, inspection of the levels of conservation of this residue at deeper phylogenetic levels suggested a potential functional impact of the change. This position is located in the boundary between the second transmembrane domain (TMD2) and the first extracellular loop (EL1) of the melanocortin-1 receptor (Figure 17). Previous works in other species have shown that mutations occurring within or near the TMD2 and TMD3 (third transmembrane domain) are mostly linked with the expression of dark colour phenotypes. These include i) the *sombre* dominant mutant phenotype in mice (Bateman, 1961), caused by either the E^{SO-3J} allele (substitution E92K) or the E^{SO} allele (substitution L98P) (Robbins *et al.*, 1993), ii) the Alaska silver phenotype of the red fox (*Vulpes vulpes*), caused by the E^{A} allele (C125R) (Våge *et al.*, 1997), iii) the black phenotype in cattle, caused by the E^{D} allele (L99P) (Klungland *et al.*, 1995), iv) the black plumage in chickens, caused by the E allele (E92K) (Takeuchi *et al.*, 1996), and v) the black coat in sheep, caused by the E^{D} allele (M73K and D121N) (Våge *et al.*, 1999). Interestingly, the L98P mutation in mice (L100P in least weasels) and the L99P mutation in cattle are located closely to the L101K identified in our work. Some of these works also tested the functional effect of these mutations and their results suggest the constitutive activation of the receptor (e.g. Robbins *et al.*, 1993; Våge *et al.*, 1997). Lu *et al.* (1998) have further tested the effect of most of these mutations in mice to propose a model that explains the constitutive activation of MC1R. Their results suggested that, for mutations tested within the two transmembrane domains [at positions E92, C123 (C125 in red foxes), and D119 (D121 in sheep)], only changes to lysine (K) and arginine (R) – positively charged basic residues – at any of these positions would result in the constitutive activation of the receptor, while simultaneously reducing the efficacy of α-MSH binding. Previous studies had

suggested the role of the acidic domain at the region of TMD2 and TMD3 (containing the negatively charged E92, D115, and D119) in the interaction with an arginine residue in the ligand, α-MSH (Haskell-Luevano *et al.*, 1996), and suggested this domain to be a part of the ligand-binding pocket (Prusis *et al.*, 1995; Haskell-Luevano *et al.*, 1996). Therefore, Lu *et al.* (1998) proposed that mutations in these regions that lead to the inclusion of a basic residue would produce the same effect in the receptor conformation as the binding to the α-MSH arginine residue, hence resulting in the constitutive activation of the receptor by ligand mimicry. They would modify the packing of TMD2 and TMD3, leading to a new conformation that would allow interaction with the G protein (Lu *et al.*, 1998; García-Borrón *et al.*, 2005). The importance of the packing of these TMDs to the activation of the receptor was further reinforced by the L98P mutation, that had the same pharmacological properties of the others (Lu *et al.*, 1998) and was predicted to modify the position of the TMDs (Robbins *et al.*, 1993).

Considering these studies, the L101K mutation identified in our work may not only cause a change in TMD2 conformation, as seen for L98P in mice (L100P in least weasels), but also the inclusion of a basic residue – lysine (K) – that could mimic the agonist binding and result in a constitutively active receptor. The constitutive activation of MC1R could further diminish the efficacy of ASIP binding to the receptor, similarly to what has been shown for α-MSH (Lu *et al.*, 1998), therefore lowering the effect of the protein as the antagonist of the receptor. Consequently, the levels of *agouti* expression that induce the change to winter-white in *nivalis* specimens (without the altered protein) could be insufficient to induce that change in *vulgaris* individuals that have the dominant mutated receptor. If so, this could explain the maintenance of the winter-brown phenotype in the latter specimens, in a mechanism similar to the one proposed for explaining the arctic fox blue phenotype (Våge *et al.*, 2005). This hypothesis could be tested by assessing the efficacy of ASIP binding to the mutated receptor. This could be accomplished, for example, through ligand binding assays similar to the ones applied by Lu *et al.* (1998) or Våge *et al.* (1999). Different ligand binding assays have been developed with recent technological advances (reviewed in de Jong *et al.*, 2005), and fluorescence-based approaches have been particularly useful for studying the interaction of ligands with G-protein coupled receptors (Sridharan *et al.*, 2014), the family of receptors where MC1R is included.

In addition, given the perfect association of the identified variants with the inheritance patterns of *nivalis-vulgaris* morphs, the *MC1R-ASIP* interaction may also be implicated in the determination of different summer colouration patterns, i.e. different demarcation lines between ventral and dorsal pelage, associated to each morph. In *ASIP*, two major transcript initiation sites are described that determine the production of two isoforms of the protein: i) the hair-cycle specific isoform, expressed in the dorsum during the mid-point of the hair growth cycle, and ii) the ventral-specific isoform, permanently expressed in the ventral part of the body

(Bultman *et al.*, 1994; Vrieling *et al.*, 1994). Differences in ventral and dorsal colouration of some mammals, particularly, lighter pigment in the ventral pelage, results from increased expression of the *agouti* ventral isoform (Bultman *et al.*, 1994; Vrieling *et al.*, 1994; Drögemüller *et al.*, 2006; Manceau *et al.*, 2011). Studies of embryonic expression of *agouti* isoforms showed that expression of the ventral-specific isoform during embryogenesis is responsible for the development of dorsal-ventral differences in colour (Millar *et al.*, 1995; Manceau *et al.*, 2011), due to inhibition of the differentiation of melanocytes in ventral epidermis (Manceau *et al.*, 2011), resulting in a demarcation line between the pelage colours. In black-spotted pigs (with black spots in a white coat background), the occurrence of *MC1R* transcripts in white areas of the coat surrounding black spots has been detected, suggesting the occurrence of melanocytes in these areas (Kijas *et al.*, 2001). Because the white coat is caused by a non-functional MC1R protein that hampers melanocyte differentiation, while black spots result from somatic reversions that re-establish MC1R functions, the authors hypothesised that detection of transcripts in some regions of white pelage could result from migration of few melanocytes from the surrounding area expressing the black phenotype. Similar migration of melanocytes at the ventral-dorsal border, coupled with the hypothesised less affinity of the *vulgaris MC1R* variant to the *agouti* antagonist, could result in different patterns of colour transition between the dorsal and ventral pelages in *nivalis* and *vulgaris* least weasels. Functional tests of the affinity of the *MC1R* variants to *agouti* and the quantification of transcripts of these genes across the boundary between the ventral and dorsal pelage in both morphs would help verifying this hypothesis, which nevertheless appears rather speculative at present.

4.3. Evidence of selection and local adaptation

Candidate genomic regions to underlie winter coat colour polymorphism identified from whole-genome scans were tested for signatures of positive selection that could be indicative of local adaptation. Results suggested the occurrence of selective sweeps in scaffolds 42 and 234 but not in scaffold 206 (Figure 18). However, only in scaffold 42 the region of the sweep coincided perfectly with the candidate region based on genome scans for differentiation. Note that the approach we applied identifies putatively selected regions based on deviations of the site frequency spectrum (SFS) from the expectations under neutrality (Pavlidis *et al.*, 2013). While poor estimates of jSFS did not allow proper demographic modelling (see section 4.1.), in this analysis larger sample sizes were used to estimate allele frequencies per phenotypic group for the selection test, which is expected to result in more accurate estimates (Buerkle and Gompert, 2013). Still, because population structure has been shown to influence inferences and parameter estimates derived from the SFS (e.g. Ptak and Przeworski, 2002;

Städler *et al.*, 2009), selection tests were replicated in the three genetic populations to verify the consistency of the results. Despite the lower sample sizes for this particular analysis, which advise caution in the interpretation of the results, inferences were perfectly consistent with the global analysis, when considering the major phenotype represented in each population (Figure S4 – Appendix IX). This suggests that the underlying genetic structure did not influence results from the selection scans.

At scaffold 234, two CLR outlier regions were identified, flanking the candidate region. In addition, while one of the putative sweeps occurs in winter-brown/*vulgaris* specimens, the other occurs in winter-white/*nivalis* individuals. These sweeps may explain the high differentiation estimates between morphs. However, its contribution to the determination of the winter coat colour polymorphism is still unclear. At scaffold 42, results suggested the occurrence of one putatively selected region in winter-brown/*vulgaris* but not in winter-white/*nivalis*, that perfectly overlaps the strongest candidate region to underlie different winter phenotypes. Interestingly, Jones *et al.* (2018) also inferred a selective sweep in the region determining winter-brown but not winter-white colour in snowshoe hares. These results suggest local adaptive pressures favouring winter-brown phenotypes. In snowshoe hares, the selective sweep in winter-brown specimens was dated to the period after the last glacial maximum, after the retreat of ice sheets (Clague and James, 2002), which would have helped the species adapt to a new environment with more transient snow (Jones *et al.*, 2018). Although we could not date the putative sweeps identified in this work, a similar scenario is plausible for the occurrence of sweeps in winter-brown least weasels in Europe. Alternatively, their occurrence may have resulted from more recent selective pressures related with environmental changes caused by global warming (Brown and Mote, 2009; Pederson *et al.*, 2011). Regardless of the time scale of the selection event, both cases point to selection favouring the brown variants during a period of climate warming and northwards retreat of the snow cover. Various studies have reported the impact of reduced or inconsistent snow cover in the fitness and distribution ranges of different species (e.g. Imperio *et al.*, 2013; Zimova *et al.*, 2016; Atmeh *et al.*, 2018); therefore, selection for the maintenance of winter-brown phenotypes could have led to adaptation to snowless environments. Because different models for future climate predict additional decreases in snow cover during this century (Danco *et al.*, 2016; Musselman *et al.*, 2017), positive selection for winter-brown phenotypes may be intensified to assure camouflage under new environments (Zimova *et al.*, 2016; Jones *et al.*, 2018).

5. Conclusions and Future Prospects

In this work, we provided the first insights into the genetic basis of seasonal coat colour variation in the least weasel (*Mustela nivalis*). Analyses of population history suggested the existence of three genetic populations in *M. nivalis* from Sweden, with partial correspondence with colour morphs, and supported the occurrence of admixture events between distinct morphs. Whole-genome scans provided an overview of the generally low genomic differentiation between morphs and allowed the identification of three highly differentiated candidate regions, and of one major candidate gene to underlie distinct winter phenotypes – *MC1R*. Additionally, screening of SNPs along this strongest candidate region allowed the identification of two variants in *MC1R* coding sequence that are candidates to cause the distinct colour phenotypes. Selective sweeps in the winter-brown/*vulgaris* colouration group suggest local adaptation favouring this phenotype, potentially driven by post-glacial climate warming and snow cover reduction.

These results are of great importance to the understanding of the evolution of seasonal coat colour polymorphism, not only in least weasels but also at a broader scale. At the physiological level, they add up to the evidence of the major role of the *MC1R-ASIP* system in the determination of colouration patterns and, particularly, of polymorphisms in winter colour across different species. At the evolutionary level, they will contribute, together with other ongoing and future projects, to paint a detailed picture of the emergence and evolution of this adaptive trait across not only the *Mustela* genus but also all mammalian species where polymorphism in winter colour is described. At the conservation level, they will facilitate the monitoring of the trait in a context of climate change, for example, by tracking changes in the distribution of allele frequencies over geographical gradients that may reflect shifts in the distributions of distinct winter phenotypes.

Future work will, on the one hand, clarify and detail the results of this work and, on the other hand, answer to newly raised questions. Even though our results were clear in identifying candidate genomic regions and loci, the precise analysis of the candidate variation associated with the different colour morphs can be further developed. First, by improving the quality and contiguity of the reference genome, building a *de novo* assembly of *Mustela nivalis* using modern assembly strategies (e.g. 10X Chromium). The *de novo* assembly of the *M. nivalis* genome would help improve mapping efficiency, recover lost SNPs, and improve the contiguity of signals that may be split into different scaffolds with the current reference genome. Second, by inferring high quality genotypes across candidate and control regions using targeted resequencing, a work we initiated here for the *MC1R* region, but at high throughput. This would

allow obtaining high-quality individual genotypes, thus improving the resolution in the strongest candidate region and clarifying the contribution of the remaining outlier regions. Third, by clarifying the apparent dissociation between winter coat colour and summer colouration type, by increasing the sample size of the *nivalis* putative winter-brown specimens with whole-genome data. Fourth, by conducting functional assays that clarify the functional impact of the candidate MC1R amino-acid change.

Also, the evolution of the winter colour polymorphism in *Mustela nivalis* remains to be fully understood. Specifically, whether the polymorphic zones where winter-white and winter-brown specimens coexist correspond to a secondary contact between diverging populations or to a transition zone of selective pressures, regardless of population structure, which would correspond to distinct evolutionary trajectories. Our results in the population from Sweden show a partial coincidence between population structure and colour morphs, which may suggest historical divergence with subsequent secondary admixture. However, a parallel analysis of the polymorphic zone from Poland, which is currently underway, will help clarifying if historical divergence between morphs explains the maintenance of distinct winter phenotypes or if this maintenance results exclusively from local selective pressures. Moreover, the study of a replicated transition zone will improve the power of the association tests.

In addition, ongoing and future work will focus on the other two species of *Mustela* that seasonally change coat colour from summer-brown to winter-white – the stoat (*M. erminea*) and the long-tailed weasel (*M. frenata*). These species, which are not sister to *M. nivalis*, are also polymorphic for winter coat colour, questioning what evolutionary mechanism created and maintained the polymorphism across the genus. Specifically, whether multiple independent evolution or a single evolutionary origin (maintained as a balanced polymorphism and/or exchanged via ancestral hybridisation) explains the convergence of the phenotypes.

Finally, the results of this work have also demonstrated the value of using natural history collections as sources of locally adapted organisms, and the feasibility of applying next-generation sequencing technologies to these samples to pinpoint candidates to explain different adaptive phenotypes, despite the technical challenges of working with museum specimens. Therefore, they encourage the use of these collections as important sources of genetic information, which can help the study of numerous species without the need to perform new invasive sampling. Applications of strategies similar to the one here described can help understand the underlying mechanisms of other important adaptive phenotypes, namely in non-model organisms that would, otherwise, be difficult or even impossible to study.

References

Aberdam, E., Bertolotto, C., Sviderskaya, E.V., de Thillot, V., Hemesath, T.J., Fisher, D.E., Bennett, D.C., Ortonne, J.-P., and Ballotti, R. (1998). Involvement of microphthalmia in the inhibition of melanocyte lineage differentiation and of melanogenesis by *agouti* signal protein. *Journal of Biological Chemistry* **273 (31)**: 19560-19565.

Abramov, A. (2000). A taxonomic review of the genus *Mustela* (Mammalia, Carnivora). *Zoosystematica Rossica* **8 (2)**: 357-364.

Abramov, A., and Baryshnikov, G. (2000). Geographic variation and intraspecific taxonomy of weasel *Mustela nivalis* (Carnivora, Mustelidae). *Zoosystematica Rossica* **8 (2)**: 365-402.

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology* **12 (2)**: R18.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19 (6)**: 716-723.

Alkan, C., Sajjadian, S., and Eichler, E.E. (2010). Limitations of next-generation genome sequence assembly. *Nature Methods* **8 (1)**: 61-65.

Alzahofi, N., Robinson, C.L., Welz, T., Page, E.L., Briggs, D.A., Stainthorp, A.K., Reekes, J., Elbe, D.A., Straub, F., Tate, E.W.*, et al.* (2018). Rab27a co-ordinates actin-dependent transport by controlling organelle-associated motors and track assembly proteins. *bioRxiv* **314153**.

Andrews, S. (2015). FastQC: a quality control tool for high throughput sequence data (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Angerbjörn, A., and Flux, J.E.C. (1995). *Lepus timidus*. *Mammalian Species* **495**: 1-11.

Atmeh, K., Andruszkiewicz, A., and Zub, K. (2018). Climate change is affecting mortality of weasels due to camouflage mismatch. *Scientific Reports* **8 (1)**: 7648.

Bailey, S.E., Mao, X.G., Struebig, M., Tsagkogeorga, G., Csorba, G., Heaney, L.R., Sedlock, J., Stanley, W., Rouillard, J.M., and Rossiter, S.J. (2016). The use of museum samples for large-scale sequence capture: a study of congeneric horseshoe bats (family Rhinolophidae). *Biological Journal of the Linnean Society* **117 (1)**: 58-70.

Bandelt, H.J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16 (1)**: 37-48.

Barlow, G.W. (1992). Is mating different in monogamous species? The Midas cichlid fish as a case study. *American Zoologist* **32 (1)**: 91-99.

Barlow, G.W., Rogers, W., and Cappeto, R.V. (1977). Incompatibility and assortative mating in the Midas cichlid. *Behavioral Ecology and Sociobiology* **2 (1)**: 49-59.

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P., and Marth, G.T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27 (12)**: 1691-1692.

Barrett, R.D.H., and Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23 (1)**: 38-44.

Barsh, G.S. (1996). The genetics of pigmentation: From fancy genes to complex traits. *Trends in Genetics* **12 (8)**: 299-305.

Bateman, N. (1961). Sombre, a viable dominant mutant in the house mouse. *Journal of Heredity* **52 (4)**: 186-189.

Beaumont, M.A., and Balding, D.J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13 (4)**: 969-980.

Bell, J. (2008). A simple way to treat PCR products prior to sequencing using ExoSAP-IT®. *BioTechniques* **44 (6)**: 834.

Bennett, D.C., and Lamoreux, M.L. (2003). The color loci of mice - a genetic century. *Pigment Cell Research* **16 (4)**: 333-344.

Bezzerides, A.L., McGraw, K.J., Parker, R.S., and Husseini, J. (2007). Elytra color as a signal of chemical defense in the Asian ladybird beetle *Harmonia axyridis*. *Behavioral Ecology and Sociobiology* **61 (9)**: 1401-1408.

Bi, K., Linderoth, T., Vanderpool, D., Good, J.M., Nielsen, R., and Moritz, C. (2013). Unlocking the vault: next-generation museum population genomics. *Molecular Ecology* **22 (24)**: 6018-6032.

Bijlsma, R., and Loeschcke, V. (2005). Environmental stress, adaptation and evolution: an overview. *Journal of Evolutionary Biology* **18 (4)**: 744-749.

Bishop, J.A. (1972). Experimental study of cline of industrial melanism in *Biston betularia* (L) (Lepidoptera) between urban Liverpool and rural North Wales. *Journal of Animal Ecology* **41 (1)**: 209-243.

Bissonnette, T.H., and Bailey, E.E. (1944). Experimental modification and control of molts and changes of coat color in weasels by controlled lighting. *Annals of the New York Academy of Sciences* **45 (6)**: 221-260.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30 (15)**: 2114-2120.

Booth, C.L. (1990). Evolutionary significance of ontogenic color-change in animals. *Biological Journal of the Linnean Society* **40 (2)**: 125-163.

Bosse, M., Spurgin, L.G., Laine, V.N., Cole, E.F., Firth, J.A., Gienapp, P., Gosler, A.G., McMahon, K., Poissant, J., Verhagen, I.*, et al.* (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science* **358 (6361)**: 365-368.

Bouzat, Juan L., Lewin, Harris A., and Paige, Ken N. (1998). The ghost of genetic diversity past: historical DNA analysis of the greater prairie chicken. *The American Naturalist* **152 (1)**: 1-6.

Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langley, C.H., and Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140 (2)**: 783-796.

Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajković, D., Kućan, Ž*., et al.* (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325 (5938)**: 318-321.

Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research* **38 (6)**: e87.

Brown, R.D., and Mote, P.W. (2009). The response of Northern Hemisphere snow cover to a changing climate. *Journal of Climate* **22 (8)**: 2124-2145.

Buerkle, C.A., and Gompert, Z. (2013). Population genomics based on low coverage sequencing: how low should we go? *Molecular Ecology* **22 (11)**: 3028-3035.

Bultman, S.J., Klebig, M.L., Michaud, E.J., Sweet, H.O., Davisson, M.T., and Woychik, R.P. (1994). Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a^t) and white-bellied agouti (A^w) reveals alternative forms of agouti transcripts. *Genes & Development* **8 (4)**: 481-490.

Burrell, A.S., Disotell, T.R., and Bergey, C.M. (2015). The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* **79**: 35-44.

Bush, W.S., and Moore, J.H. (2012). Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology* **8 (12)**: e1002822.

Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C*., et al.* (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics* **43 (10)**: 956-963.

Caprioli, M., Ambrosini, R., Boncoraglio, G., Gatti, E., Romano, A., Romano, M., Rubolini, D., Gianfranceschi, L., and Saino, N. (2012). Clock gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. *PLoS ONE* **7 (4)**: e35140.

Cavazza, F. (1909). Studien über die in Italien vorkommenden Wieselarten der Untergattung Arctogale. *Zoologischer Anzeiger* **34**: 582-603.

Chen, C.-F., Foley, J., Tang, P.-C., Li, A., Jiang, T.X., Wu, P., Widelitz, R.B., and Chuong, C.M. (2015). Development, regeneration, and evolution of feathers. *Annual Review of Animal Biosciences* **3 (1)**: 169-195.

Clague, J.J., and James, T.S. (2002). History and isostatic effects of the last ice sheet in southern British Columbia. *Quaternary Science Reviews* **21 (1)**: 71-87.

Corbet, G.B. (1978). *The mammals of the Palaearctic region: a taxonomic review*. London: British Museum (Natural History); Ithaca (N.Y.): Cornell University Press.

Corbet, G.B., and Hill, J.E. (1991). *A world list of mammalian species*, 3 ed. London: British Museum (Natural History).

Crawford, J.E., and Lazzaro, B. (2012). Assessing the accuracy and power of population genetic inference from low-pass next-generation sequencing data. *Frontiers in Genetics* **3 (66)**: 1-13.

Crawford, J.E., and Nielsen, R. (2013). Detecting adaptive trait loci in nonmodel systems: divergence or admixture mapping? *Molecular Ecology* **22 (24)**: 6131-6148.

D'Andrea, A.D., and Grompe, M. (2003). The Fanconi anaemia/BRCA pathway. *Nature Reviews Cancer* **3**: 23-34.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.L.*, et al.* (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* **110 (39)**: 15758-15763.

Dabney, J., and Meyer, M. (2012). Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques* **52 (2)**: 87-94.

Danco, J.F., DeAngelis, A.M., Raney, B.K., and Broccoli, A.J. (2016). Effects of a warming climate on daily snowfall events in the Northern Hemisphere. *Journal of Climate* **29 (17)**: 6295-6318.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**: 499-510.

de Jong, L.A.A., Uges, D.R.A., Franke, J.P., and Bischoff, R. (2005). Receptor–ligand binding assays: technologies and applications. *Journal of Chromatography B* **829 (1)**: 1-25.

Degerbøl, M. (1940). Mammalia. In *The Zoology of the Faroes*. Copenhagen: University Zoological Museum.

Dietrich, S., Weiß, S., Pleiser, S., and Kerkhoff, E. (2013). Structural and functional insights into the Spir/formin actin nucleator complex. *Biological chemistry* **394 (12)**: 1649-1660.

Drögemüller, C., Giese, A., Martins-Wess, F., Wiedemann, S., Andersson, L., Brenig, B., Fries, R., and Leeb, T. (2006). The mutation causing the black-and-tan pigmentation phenotype of Mangalitza pigs maps to the porcine *ASIP* locus but does not affect its coding sequence. *Mammalian Genome* **17 (1)**: 58-66.

Du Merle, P. (1999). Egg development and diapause: ecophysiological and genetic basis of phenological polymorphism and adaptation to varied hosts in the green oak tortrix, *Tortrix viridana* L. (Lepidoptera: Tortricidae). *Journal of Insect Physiology* **45 (6)**: 599-611.

Duncan, M.J., and Goldman, B.D. (1984). Hormonal regulation of the annual pelage color cycle in the Djungarian hamster, *Phodopus sungorus*. I. Role of the gonads and the pituitary. *Journal of Experimental Zoology* **230 (1)**: 89-95.

Easterla, D.A. (1970). First records of the least weasel, *Mustela nivalis*, from Missouri and southwestern Iowa. *Journal of Mammalogy* **51 (2)**: 333-340.

Ehrlich, J., Sankoff, D., and Nadeau, J.H. (1997). Synteny conservation and chromosome rearrangements during mammalian evolution. *Genetics* **147 (1)**: 289-296.

Elzinga, J.A., Atlan, A., Biere, A., Gigord, L., Weis, A.E., and Bernasconi, G. (2007). Time after time: flowering phenology and biotic interactions. *Trends in Ecology & Evolution* **22 (8)**: 432-439.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14 (8)**: 2611-2620.

Evans, Richard D., Robinson, C., Briggs, Deborah A., Tooth, David J., Ramalho, Jose S., Cantero, M., Montoliu, L., Patel, S., Sviderskaya, Elena V., and Hume, Alistair N. (2014). Myosin-Va and dynamic actin oppose microtubules to drive long-range organelle transport. *Current Biology* **24 (15)**: 1743-1750.

Everts, R.E., Rothuizen, J., and Oost, B.A. (2000). Identification of a premature stop codon in the melanocyte-stimulating hormone receptor gene (MC1R) in Labrador and Golden retrievers with yellow coat colour. *Animal Genetics* **31 (3)**: 194-199.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., and Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics* **9 (10)**: e1003905.

Excoffier, L., Hofer, T., and Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity* **103**: 285-298.

Excoffier, L., and Lischer, H.E.L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10 (3)**: 564-567.

Fabre, P.-H., Vilstrup, J.T., Raghavan, M., Der Sarkissian, C., Willerslev, E., Douzery, E.J.P., and Orlando, L. (2014). Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology Letters* **10 (7)**: 1-5.

Fay, J.C., and Wu, C.-I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* **155 (3)**: 1405-1413.

Ferreira, M.S., Alves, P.C., Callahan, C.M., Marques, J.P., Mills, L.S., Good, J.M., and Melo-Ferreira, J. (2017). The transcriptional landscape of seasonal coat colour moult in the snowshoe hare. *Molecular Ecology* **26 (16)**: 4173-4185.

Fisher, R.A. (1922). On the interpretation of χ2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society* **85 (1)**: 87-94.

Flux, J.E.C. (1970). Colour change of mountain hares (*Lepus timidus scoticus*) in north-east Scotland. *Journal of Zoology* **162 (3)**: 345-358.

Flux, J.E.C., and Angermann, R. (1990). The Hares and Jackrabbits. In *Rabbits, Hares and Pikas: Status Survey and Conservation Action Plan*. J.A. Chapman, and J.E.C. Flux, eds. Gland, Switzerland: IUCN/SSC Lagomorph Specialist Group.

Fontanesi, L., Tazzoli, M., Beretti, F., and Russo, V. (2006). Mutations in the melanocortin 1 receptor (MC1R) gene are associated with coat colours in the domestic rabbit (*Oryctolagus cuniculus*). *Animal Genetics* **37 (5)**: 489-493.

Forrest, J., and Miller-Rushing, A.J. (2010). Toward a synthetic understanding of the role of phenology in ecology and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365 (1555)**: 3101-3112.

Frank, F. (1985). On the evolution and systematics of the weasels (*Mustela nivalis* L.). *Zeitschrift für Säugetierkunde* **50 (4)**: 208-225.

García-Borrón, J.C., Sánchez-Laorden, B.L., and Jiménez-Cervantes, C. (2005). Melanocortin-1 receptor structure and functional regulation. *Pigment Cell Research* **18 (6)**: 393-410.

Goldman, B.D. (2001). Mammalian photoperiodic system: formal properties and neuroendocrine mechanisms of photoperiodic time measurement. *Journal of Biological Rhythms* **16 (4)**: 283-301.

Gower, B.A., Nagy, T.R., and Stetson, M.H. (1993). Role of prolactin and the gonads in seasonal physiological changes in the collared lemming (*Dicrostonyx groenlandicus*). *Journal of Experimental Zoology* **266 (2)**: 92-101.

Grant, P.R., and Grant, B.R. (2006). Evolution of character displacement in Darwin's finches. *Science* **313 (5784)**: 224-226.

Guillermo-Ferreira, R., Bispo, P.C., Appel, E., Kovalev, A., and Gorb, S.N. (2015). Mechanism of the wing colouration in the dragonfly *Zenithoptera lanei* (Odonata: Libellulidae) and its role in intraspecific communication. *Journal of Insect Physiology* **81**: 129-136.

Guschanski, K., Krause, J., Sawyer, S., Valente, L.M., Bailey, S., Finstermeier, K., Sabin, R., Gilissen, E., Sonet, G., Nagy, Z.T.*, et al.* (2013). Next-generation museomics disentangles one of the largest primate radiations. *Systematic Biology* **62 (4)**: 539-554.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., and Bustamante, C.D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* **5 (10)**: e1000695.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95-98.

Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., Schork, N.J., Murray, S.S., Topol, E.J., Levy, S.*, et al.* (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology* **10 (3)**: R32.

Haskell-Luevano, C., Sawyer, T.K., Trumpp-Kallmeyer, S., Bikker, J.A., Humblet, C., Gantz, I., and Hruby, V.J. (1996). Three-dimensional molecular models of the hMC1R melanocortin receptor: complexes with melanotropin peptide agonists. *Drug Design and Discovery* **14 (3)**: 197-211.

Hawkins, R.D., Hon, G.C., and Ren, B. (2010). Next-generation genomics: an integrative approach. *Nature Reviews Genetics* **11**: 476-486.

Hearing, V.J., and Tsukamoto, K. (1991). Enzymatic control of pigmentation in mammals. *The FASEB Journal* **5 (14)**: 2902-2909.

Hernandez, R.D., Hubisz, M.J., Wheeler, D.A., Smith, D.G., Ferguson, B., Rogers, J., Nazareth, L., Indap, A., Bourquin, T., McPherson, J.*, et al.* (2007). Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. *Science* **316 (5822)**: 240-243.

Hewson, R. (1973). The moults of captive Scottish ptarmigan (*Lagopus mutus*). *Journal of Zoology* **171 (2)**: 177-187.

Hewson, R., and Watson, A. (1979). Winter whitening of stoats (*Mustela erminea*) in Scotland and north-east England. *Journal of Zoology* **187 (1)**: 55-64.

Hoekstra, H.E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **97 (3)**: 222-234.

Hoekstra, H.E., Hirschmann, R.J., Bundey, R.A., Insel, P.A., and Crossland, J.P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313 (5783)**: 101-104.

Hoffmann, A.A., and Sgrò, C.M. (2011). Climate change and evolutionary adaptation. *Nature* **470**: 479-485.

Hofman, M.A. (2004). The brain's calendar: neural mechanisms of seasonal timing. *Biological Reviews* **79 (1)**: 61-77.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., and Cresko, W.A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6 (2)**: e1000862.

Huang, X.H., Zhao, Y., Wei, X.H., Li, C.Y., Wang, A., Zhao, Q., Li, W.J., Guo, Y.L., Deng, L.W., Zhu, C.R*., et al.* (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* **44**: 32-39.

Hubbard, J.K., Uy, J.A.C., Hauber, M.E., Hoekstra, H.E., and Safran, R.J. (2010). Vertebrate pigmentation: from underlying genes to adaptive function. *Trends in Genetics* **26 (5)**: 231-239.

Imperio, S., Bionda, R., Viterbi, R., and Provenzale, A. (2013). Climate change and human disturbance can lead to local extinction of Alpine rock ptarmigan: new insight from the Western Italian Alps. *PLoS ONE* **8 (11)**: e81598.

Jacobsen, E.E., White, C.M., and Emison, W.B. (1983). Molting adaptations of rock ptarmigan on Amchitka Island, Alaska. *Condor* **85 (4)**: 420-426.

Jones, M.R., Mills, L.S., Alves, P.C., Callahan, C.M., Alves, J.M., Lafferty, D.J.R., Jiggins, F.M., Jensen, J.D., Melo-Ferreira, J., and Good, J.M. (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360 (6395)**: 1355-1358.

Kaplan, N.L., Hudson, R.R., and Langley, C.H. (1989). The "hitchhiking effect" revisited. *Genetics* **123 (4)**: 887-899.

Karlsson, E.K., Baranowska, I., Wade, C.M., Salmon Hillbertz, N.H.C., Zody, M.C., Anderson, N., Biagi, T.M., Patterson, N., Pielberg, G.R., Kulbokas Iii, E.J*., et al.* (2007). Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature Genetics* **39**: 1321-1328.

Kerkhoff, E., Simpson, J.C., Leberfinger, C.B., Otto, I.M., Doerks, T., Bork, P., Rapp, U.R., Raabe, T., and Pepperkok, R. (2001). The Spir actin organizers are involved in vesicle transport processes. *Current Biology* **11 (24)**: 1963-1968.

Kijas, J.M.H., Moller, M., Plastow, G., and Andersson, L. (2001). A frameshift mutation in *MC1R* and a high frequency of somatic reversions cause black spotting in pigs. *Genetics* **158 (2)**: 779-785.

Kijas, J.M.H., Wales, R., Tornsten, A., Chardon, P., Moller, M., and Andersson, L. (1998). Melanocortin receptor 1 (*MC1R*) mutations and coat color in pigs. *Genetics* **150 (3)**: 1177-1185.

Kim, S.Y., Lohmueller, K.E., Albrechtsen, A., Li, Y., Korneliussen, T., Tian, G., Grarup, N., Jiang, T., Andersen, G., Witte, D*., et al.* (2011). Estimation of allele frequency and

association mapping using next-generation sequencing data. *BMC Bioinformatics* **12 (1)**: 231.

King, C.M. (1979). Molt and color change in english weasels (*Mustela nivalis*). *Journal of Zoology* **189 (2)**: 127-134.

King, C.M., and Powell, R.A. (2007). *The natural history of weasels and stoats: ecology, behavior, and management*. New York: Oxford University Press, Inc.

Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C.E., Hoang, A., Gibert, P., and Beerli, P. (2001). The strength of phenotypic selection in natural populations. *The American Naturalist* **157 (3)**: 245-261.

Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research* **40 (1)**: e3.

Klungland, H., Vage, D.I., Gomez-Raya, L., Adalsteinsson, S., and Lien, S. (1995). The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mammalian Genome* **6 (9)**: 636-639.

Ko, M.S., Lee, U.H., Kim, S.I., Kim, H.J., Park, J.J., Cha, S.J., Kim, S.B., Song, H., Chung, D.K., Han, I.S.*, et al.* (2004). Overexpression of DRG2 suppresses the growth of Jurkat T cells but does not induce apoptosis. *Archives of Biochemistry and Biophysics* **422 (2)**: 137-144.

Koepfli, K.-P., Deere, K.A., Slater, G.J., Begg, C., Begg, K., Grassman, L., Lucherini, M., Veron, G., and Wayne, R.K. (2008). Multigene phylogeny of the Mustelidae: resolving relationships, tempo and biogeographic history of a mammalian adaptive radiation. *BMC Biology* **6 (1)**: 10.

Kofler, R., Pandey, R.V., and Schlotterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27 (24)**: 3435-3436.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15 (5)**: 1179-1191.

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15 (1)**: 356.

Kumar, S., and Subramanian, S. (2002). Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences* **99 (2)**: 803-808.

Kurose, N., Abramov, A.V., and Masuda, R. (2000). Intrageneric diversity of the cytochrome b gene and phylogeny of Eurasian species of the genus *Mustela* (Mustelidae, Carnivora). *Zoological Science* **17 (5)**: 673-679.

Kurose, N., Abramov, A.V., and Masuda, R. (2008). Molecular phylogeny and taxonomy of the genus *Mustela* (Mustelidae, Carnivora), inferred from mitochondrial DNA sequences: new perspectives on phylogenetic status of the back-striped weasel and American mink. *Mammal Study* **33 (1)**: 25-33.

Laity, J.H., Lee, B.M., and Wright, P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology* **11 (1)**: 39-46.

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E*., et al.* (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307 (5715)**: 1618-1621.

Le Pape, E., Passeron, T., Giubellino, A., Valencia, J.C., Wolber, R., and Hearing, V.J. (2009). Microarray analysis sheds light on the dedifferentiating role of agouti signal protein in murine melanocytes via the Mc1r. *Proceedings of the National Academy of Sciences* **106 (6)**: 1802-1807.

Le, S.Q., and Durbin, R. (2011). SNP detection and genotyping from low-coverage sequencing data on multiple diploid samples. *Genome Research* **21 (6)**: 952-960.

Lebarbenchon, C., Poitevin, F., Arnal, V., and Montgelard, C. (2010). Phylogeography of the weasel (*Mustela nivalis*) in the western-Palaearctic region: combined effects of glacial events and human movements. *Heredity* **105 (5)**: 449-462.

Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30 (20)**: 2843-2851.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25 (14)**: 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25 (16)**: 2078-2079.

Li, Y., Sidore, C., Kang, H.M., Boehnke, M., and Abecasis, G.R. (2011). Low-coverage sequencing: implications for design of complex trait association studies. *Genome Research* **21 (6)**: 940-951.

Lim, H.C., and Braun, M.J. (2016). High-throughput SNP genotyping of historical and modern samples of five bird species via sequence capture of ultraconserved elements. *Molecular Ecology Resources* **16 (5)**: 1204-1223.

Lincoln, G.A., Clarke, I.J., Hut, R.A., and Hazlerigg, D.G. (2006). Characterizing a mammalian circannual pacemaker. *Science* **314 (5807)**: 1941-1944.

Linn, I., and Day, M.G. (1966). Identification of individual weasels *Mustela nivalis* using ventral pelage pattern. *Journal of Zoology* **148 (4)**: 583-585.

Liu, Q., Liu, W., Zeng, B., Wang, G., Hao, D., and Huang, Y. (2017). Deletion of the *Bombyx mori* odorant receptor co-receptor (BmOrco) impairs olfactory sensitivity in silkworms. *Insect Biochemistry and Molecular Biology* **86**: 58-67.

Logan, A., and Weatherhead, B. (1978). Pelage color cycles and hair follicle tyrosinase activity in siberian hamster. *Journal of Investigative Dermatology* **71 (5)**: 295-298.

Lopes, R.J., Johnson, J.D., Toomey, M.B., Ferreira, M.S., Araujo, P.M., Melo-Ferreira, J., Andersson, L., Hill, G.E., Corbo, J.C., and Carneiro, M. (2016). Genetic basis for red coloration in birds. *Current Biology* **26 (11)**: 1427-1434.

Lu, D.S., Våge, D.I., and Cone, R.D. (1998). A ligand-mimetic model for constitutive activation of the melanocortin-1 receptor. *Molecular Endocrinology* **12 (4)**: 592-604.

Lu, D.S., Willard, D., Patel, I.R., Kadwell, S., Overton, L., Kost, T., Luther, M., Chen, W.B., Woychik, R.P., Wilkison, W.O.*, et al.* (1994). Agouti protein is an antagonist of the melanocyte-stimulating hormone receptor. *Nature* **371 (6500)**: 799-802.

Malek, T.B., Boughman, J.W., Dworkin, I., and Peichel, C.L. (2012). Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Molecular Ecology* **21 (21)**: 5265-5279.

Manceau, M., Domingues, V.S., Mallarino, R., and Hoekstra, H.E. (2011). The developmental role of Agouti in color pattern evolution. *Science* **331 (6020)**: 1062-1065.

Mardis, E.R. (2011). A decade's perspective on DNA sequencing technology. *Nature* **470**: 198-203.

McCormack, J.E., Tsai, W.L.E., and Faircloth, B.C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources* **16 (5)**: 1189-1203.

McDevitt, A.D., Zub, K., Kawalko, A., Oliver, M.K., Herman, J.S., and Wojcik, J.M. (2012). Climate and refugial origin influence the mitochondrial lineage distribution of weasels (*Mustela nivalis*) in a phylogeographic suture zone. *Biological Journal of the Linnean Society* **106 (1)**: 57-69.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.*, et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20 (9)**: 1297-1303.

Mercer, J.A., Seperack, P.K., Strobel, M.C., Copeland, N.G., and Jenkins, N.A. (1991). Novel myosin heavy chain encoded by murine dilute coat colour locus. *Nature* **349**: 709-713.

Metzker, M.L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**: 31-46.

Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010 (6)**: pdb.prot5448.

Miles, C., and Wayne, M. (2008). Quantitative trait locus (QTL) analysis. *Nature Education* **1 (1)**: 208.

Millar, S.E., Miller, M.W., Stevens, M.E., and Barsh, G.S. (1995). Expression and transgenic studies of the mouse agouti gene provide insight into the mechanisms by which mammalian coat color patterns are generated. *Development* **121 (10)**: 3223-3232.

Miller, W., Drautz, D.I., Janecka, J.E., Lesk, A.M., Ratan, A., Tomsho, L.P., Packard, M., Zhang, Y., McClellan, L.R., Qi, J.*, et al.* (2009). The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*). *Genome Research* **19 (2)**: 213-220.

Mills, L.S., Bragina, E.V., Kumar, A.V., Zimova, M., Lafferty, D.J.R., Feltner, J., Davis, B.M., Hackländer, K., Alves, P.C., Good, J.M.*, et al.* (2018). Winter color polymorphisms identify global hot spots for evolutionary rescue from climate change. *Science* **359 (6379)**: 1033-1036.

Mills, L.S., Zimova, M., Oyler, J., Running, S., Abatzoglou, J.T., and Lukacs, P.M. (2013). Camouflage mismatch in seasonal coat color due to decreased snow duration. *Proceedings of the National Academy of Sciences* **110 (18)**: 7360-7365.

Montgomerie, R., Lyon, B., and Holder, K. (2001). Dirty ptarmigan: behavioral modification of conspicuous male plumage. *Behavioral Ecology* **12 (4)**: 429-438.

Moore, R.Y. (1995). Neural control of the pineal gland. *Behavioural Brain Research* **73 (1)**: 125-130.

Moritz, C., Dowling, T.E., and Brown, W.M. (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* **18 (1)**: 269-292.

Musselman, K.N., Clark, M.P., Liu, C., Ikeda, K., and Rasmussen, R. (2017). Slower snowmelt in a warmer world. *Nature Climate Change* **7**: 214-219.

Nachman, M.W. (2013). Genomics and museum specimens. *Molecular Ecology* **22 (24)**: 5966-5968.

Nachman, M.W., Hoekstra, H.E., and D'Agostino, S.L. (2003). The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences* **100 (9)**: 5268-5273.

Nadeau, N.J., Whibley, A., Jones, R.T., Davey, J.W., Dasmahapatra, K.K., Baxter, S.W., Quail, M.A., Joron, M., ffrench-Constant, R.H., Blaxter, M.L.*, et al.* (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367 (1587)**: 343-353.

Nagorsen, D.W. (1983). Winter pelage colour in snowshoe hares (*Lepus americanus*) from the Pacific Northwest. *Canadian Journal of Zoology* **61 (10)**: 2313-2318.

Ng, P.C., and Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31 (13)**: 3812-3814.

Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A.*, et al.* (2009). Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics* **42**: 30-35.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics* **39 (1)**: 197-218.

Nielsen, R., Hubisz, M.J., Torgerson, D., Andres, A.M., Albrechtsen, A., Gutenkunst, R., Adams, M., Cargill, M., Boyko, A., and Indap, A. (2009). Darwinian and demographic forces affecting human protein coding genes. *Genome Research* **19 (5)**: 838-849.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., and Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS ONE* **7 (7)**: e37558.

Nielsen, R., Paul, J.S., Albrechtsen, A., and Song, Y.S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* **12**: 443-451.

Nunome, M., Kinoshita, G., Tomozawa, M., Torii, H., Matsuki, R., Yamada, F., Matsuda, Y., and Suzuki, H. (2014). Lack of association between winter coat colour and genetic

population structure in the Japanese hare, *Lepus brachyurus* (Lagomorpha: Leporidae). *Biological Journal of the Linnean Society* **111 (4)**: 761-776.

Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32 (2)**: 292-294.

Oleksyk, T.K., Smith, M.W., and O'Brien, S.J. (2010). Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365 (1537)**: 185-205.

Olsson, M., Durbeej, M., Ekblom, P., and Hjalt, T. (2002). Nulp1, a novel basic helix-loop-helix protein expressed broadly during early embryonic organogenesis and prominently in developing dorsal root ganglia. *Cell and Tissue Research* **308 (3)**: 361-370.

Otto, E.A., Hurd, T.W., Airik, R., Chaki, M., Zhou, W., Stoetzel, C., Patil, S.B., Levy, S., Ghosh, A.K., Murga-Zamalloa, C.A*., et al.* (2010). Candidate exome capture identifies mutation of *SDCCAG8* as the cause of a retinal-renal ciliopathy. *Nature Genetics* **42**: 840-850.

Papa, R., Kapan, D.D., Counterman, B.A., Maldonado, K., Lindstrom, D.P., Reed, R.D., Nijhout, H.F., Hrbek, T., and McMillan, W.O. (2013). Multi-allelic major effect genes interact with minor effect QTLs to control adaptive color pattern variation in *Heliconius erato*. *PLoS ONE* **8 (3)**: e57033.

Parchem, R.J., Perry, M.W., and Patel, N.H. (2007). Patterns on the insect wing. *Current Opinion in Genetics & Development* **17 (4)**: 300-308.

Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W., and Buerkle, C.A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* **21 (12)**: 2991-3005.

Pasaniuc, B., Rohland, N., McLaren, P.J., Garimella, K., Zaitlen, N., Li, H., Gupta, N., Neale, B.M., Daly, M.J., Sklar, P*., et al.* (2012). Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nature Genetics* **44**: 631-635.

Pauli, J.N., Zuckerberg, B., Whiteman, J.P., and Porter, W. (2013). The subnivium: a deteriorating seasonal refugium. *Frontiers in Ecology and the Environment* **11 (5)**: 260-267.

Pavlidis, P., Zivkovic, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution* **30 (9)**: 2224-2234.

Payne, R.B., and Soreson, M.D. (2002). Museum collections as sources of genetic data. *Bonner zoologische Beiträge* **51 (2/3)**: 97-104.

Pederson, G.T., Gray, S.T., Woodhouse, C.A., Betancourt, J.L., Fagre, D.B., Littell, J.S., Watson, E., Luckman, B.H., and Graumlich, L.J. (2011). The unusual nature of recent snowpack declines in the North American Cordillera. *Science* **333 (6040)**: 332-335.

Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8 (11)**: e1002967.

Pool, J., Hellmann, I., Jensen, J., and Nielsen, R. (2010). Population genetics of genome-scale sequence variation. *Genome Research* **20 (3)**: 291-300.

Poulakakis, N., Glaberman, S., Russello, M., Beheregaray, L.B., Ciofi, C., Powell, J.R., and Caccone, A. (2008). Historical DNA analysis reveals living descendants of an extinct

species of Galapagos tortoise. *Proceedings of the National Academy of Sciences* **105 (40)**: 15464-15469.

Price, A.L., Patterson, N., Yu, F., Cox, D.R., Waliszewska, A., McDonald, G.J., Tandon, A., Schirmer, C., Neubauer, J., Bedoya, G.*, et al.* (2007). A genomewide admixture map for latino populations. *The American Journal of Human Genetics* **80 (6)**: 1024-1036.

Pritchard, J.K., Stephens, M., Rosenberg, N.A., and Donnelly, P. (2000). Association mapping in structured populations. *The American Journal of Human Genetics* **67 (1)**: 170-181.

Protas, M.E., and Patel, N.H. (2008). Evolution of coloration patterns. *Annual Review of Cell and Developmental Biology* **24 (1)**: 425-446.

Provance, D.W., Wei, M., Ipe, V., and Mercer, J.A. (1996). Cultured melanocytes from dilute mutant mice exhibit dendritic morphology and altered melanosome distribution. *Proceedings of the National Academy of Sciences* **93 (25)**: 14554-14558.

Prusis, P., Frandberg, P.A., Muceniece, R., Kalvinsh, I., and Wikberg, J.E.S. (1995). A three-dimensional model for the interaction of MSH with the melanocortin-1 receptor. *Biochemical and Biophysical Research Communications* **210 (1)**: 205-210.

Pryke, S.R., and Griffith, S.C. (2007). The relative role of male vs. female mate choice in maintaining assortative pairing among discrete colour morphs. *Journal of Evolutionary Biology* **20 (4)**: 1512-1521.

Ptak, S.E., and Przeworski, M. (2002). Evidence for population growth in humans is confounded by fine-scale population structure. *Trends in Genetics* **18 (11)**: 559-563.

Purdue, J.R., Smith, M.H., and Patton, J.C. (2000). Female philopatry and extreme spatial genetic heterogeneity in white-tailed deer. *Journal of Mammalogy* **81 (1)**: 179-185.

Rabosky, A.R.D., Cox, C.L., Rabosky, D.L., Title, P.O., Holmes, I.A., Feldman, A., and McGuire, J.A. (2016). Coral snakes predict the evolution of mimicry across New World snakes. *Nature Communications* **7**: 11484.

R Core Team (2016). R: A language and environment for statistical computing (https://www.r-project.org/). Vienna, Austria: R Foundation for Statistical Computing.

Robbins, L.S., Nadeau, J.H., Johnson, K.R., Kelly, M.A., Rosellirehfuss, L., Baack, E., Mountjoy, K.G., and Cone, R.D. (1993). Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell* **72 (6)**: 827-834.

Rodrigues, M.A.F. (2015). Phylogeography and evolutionary genetics of the weasel (*Mustela nivalis*). PhD thesis. University of Lisbon, Portugal.

Rothschild, M. (1942). Change of pelage in the stoat *Mustela erminea* L. *Nature* **149**: 78.

Roulin, A. (2004). The evolution, maintenance and adaptive function of genetic colour polymorphism in birds. *Biological Reviews* **79 (4)**: 815-848.

Rouzaud, F., Annereau, J.P., Valencia, J.C., Costin, G.E., and Hearing, V.J. (2003). Regulation of melanocortin 1 receptor expression at the mRNA and protein levels by its natural agonist and antagonist. *The FASEB Journal* **17 (12)**: 2154-2156.

Rowe, K.C., Singhal, S., Macmanes, M.D., Ayroles, J.F., Morelli, T.L., Rubidge, E.M., Bi, K., and Moritz, C.C. (2011). Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources* **11 (6)**: 1082-1092.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., and Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* **34 (12)**: 3299-3302.

Rust, C.C. (1962). Temperature as a modifying factor in the spring pelage change of short-tailed weasels. *Journal of Mammalogy* **43 (3)**: 323-328.

Rust, C.C. (1965). Hormonal control of pelage cycles in short tailed weasel (*Mustela erminea bangsi*). *General and Comparative Endocrinology* **5 (2)**: 222-231.

Rust, C.C., and Meyer, R.K. (1969). Hair color, molt, and testis size in male, short-tailed weasels treated with melatonin. *Science* **165 (3896)**: 921-922.

Santer, B.D., Po-Chedley, S., Zelinka, M.D., Cvijanovic, I., Bonfils, C., Durack, P.J., Fu, Q., Kiehl, J., Mears, C., Painter, J.*, et al.* (2018). Human influence on the seasonal cycle of tropospheric temperature. *Science* **361 (6399)**: eaas8806.

Sarver, B.A.J., Keeble, S., Cosart, T., Tucker, P.K., Dean, M.D., and Good, J.M. (2017). Phylogenomic insights into mouse evolution using a pseudoreference approach. *Genome Biology and Evolution* **9 (3)**: 726-739.

Sato, J.J., Hosoda, T., Wolsan, M., Tsuchiya, K., Yamamoto, M., and Suzuki, H. (2003). Phylogenetic relationships and divergence times among Mustelids (Mammalia: Carnivora) based on nucleotide sequences of the nuclear interphotoreceptor retinoid binding protein and mitochondrial cytochrome b genes. *Zoological Science* **20 (2)**: 243-264.

Sato, J.J., Wolsan, M., Prevosti, F.J., D'Elía, G., Begg, C., Begg, K., Hosoda, T., Campbell, K.L., and Suzuki, H. (2012). Evolutionary and biogeographic history of weasel-like carnivorans (Musteloidea). *Molecular Phylogenetics and Evolution* **63 (3)**: 745-757.

Schlötterer, C., Tobler, R., Kofler, R., and Nolte, V. (2014). Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* **15**: 749-763.

Schneider, A., David, V.A., Johnson, W.E., O'Brien, S.J., Barsh, G.S., Menotti-Raymond, M., and Eizirik, E. (2012). How the leopard hides its spots: *ASIP* mutations and melanism in wild cats. *PLoS ONE* **7 (12)**: e50386.

Schuh, M. (2011). An actin-dependent mechanism for long-range vesicle transport. *Nature Cell Biology* **13**: 1431-1436.

Seixas, F.A., Boursot, P., and Melo-Ferreira, J. (2018). The genomic impact of historical hybridization with massive mitochondrial DNA introgression. *Genome Biology* **19 (1)**: 91.

Severaid, J.H. (1945). Pelage changes in the snowshoe hare (*Lepus americanus struthopus* Bangs). *Journal of Mammalogy* **26 (1)**: 41-63.

Shapiro, B., Sibthorpe, D., Rambaut, A., Austin, J., Wragg, G.M., Bininda-Emonds, O.R.P., Lee, P.L.M., and Cooper, A. (2002). Flight of the dodo. *Science* **295 (5560)**: 1683-1683.

Sheffield, S.R., and King, C.M. (1994). *Mustela nivalis*. *Mammalian Species* **454**: 1-10.

Silva, C., Besnard, G., Piot, A., Razanatsoa, J., Oliveira, R.P., and Vorontsova, M.S. (2017). Museomics resolve the systematics of an endangered grass lineage endemic to north-western Madagascar. *Annals of Botany* **119 (3)**: 339-351.

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research* **40 (W1)**: W452-W457.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Research* **19 (6)**: 1117-1123.

Skoglund, P., and Hoglund, J. (2010). Sequence polymorphism in candidate genes for differences in winter plumage between Scottish and Scandinavian willow grouse (*Lagopus lagopus*). *PLoS ONE* **5 (4)**: e10334.

Skotte, L., Korneliussen, T.S., and Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195 (3)**: 693-702.

Slominski, A., Tobin, D.J., Shibahara, S., and Wortsman, J. (2004). Melanin pigmentation in mammalian skin and its hormonal regulation. *Physiological Reviews* **84 (4)**: 1155-1228.

Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23 (1)**: 23-35.

Solomon, S., Plattner, G.-K., Knutti, R., and Friedlingstein, P. (2009). Irreversible climate change due to carbon dioxide emissions. *Proceedings of the National Academy of Sciences* **106 (6)**: 1704-1709.

Speed, H.E., Kouser, M., Xuan, Z., Reimers, J.M., Ochoa, C.F., Gupta, N., Liu, S., and Powell, C.M. (2015). Autism-associated insertion mutation (InsG) of *Shank3* exon 21 causes impaired synaptic transmission and behavioral deficits. *The Journal of Neuroscience* **35 (26)**: 9648-9665.

Sridharan, R., Zuber, J., Connelly, S.M., Mathew, E., and Dumont, M.E. (2014). Fluorescent approaches for understanding interactions of ligands with G protein coupled receptors. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1838 (1, Part A)**: 15-33.

Staats, M., Erkens, R.H.J., van de Vossenberg, B., Wieringa, J.J., Kraaijeveld, K., Stielow, B., Geml, J., Richardson, J.E., and Bakker, F.T. (2013). Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS ONE* **8 (7)**: e69189.

Städler, T., Haubold, B., Merino, C., Stephan, W., and Pfaffelhuber, P. (2009). The impact of sampling schemes on the site frequency spectrum in non-equilibrium subdivided populations. *Genetics* **182**: 205-216.

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., and Slate, J. (2010). Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25 (12)**: 705-712.

Steen, J.B., Erikstad, K.E., xf, and idal, K. (1992). Cryptic behaviour in moulting hen willow ptarmigan *Lagopus l. lagopus* during snow melt. *Ornis Scandinavica (Scandinavian Journal of Ornithology)* **23 (1)**: 101-104.

Stenn, K.S., and Paus, R. (2001). Controls of hair follicle cycling. *Physiological Reviews* **81 (1)**: 449-494.

Stolt, B.O. (1979). Color pattern and size variation of the weasel *Mustela nivalis* L. in Sweden. *Zoon* **7 (1)**: 55-61.

Stuart-Fox, D., and Moussalli, A. (2009). Camouflage, communication and thermoregulation: lessons from colour changing organisms. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364 (1516)**: 463-470.

Suarez, A.V., and Tsutsui, N.D. (2004). The value of museum collections for research and society. *BioScience* **54 (1)**: 66-74.

Sultaire, S.M., Pauli, J.N., Martin, K.J., Meyer, M.W., Notaro, M., and Zuckerberg, B. (2016). Climate change surpasses land-use change in the contracting range boundary of a winter-adapted mammal. *Proceedings of the Royal Society B: Biological Sciences* **283 (1827)**: 20153104.

Summers, K., and Clough, M.E. (2001). The evolution of coloration and toxicity in the poison frog family (Dendrobatidae). *Proceedings of the National Academy of Sciences* **98 (11)**: 6227-6232.

Sviderskaya, E.V., Hill, S.P., Balachandar, D., Barsh, G.S., and Bennett, D.C. (2001). Agouti signaling protein and other factors modulating differentiation and proliferation of immortal melanoblasts. *Developmental Dynamics* **221 (4)**: 373-379.

Takeuchi, S., Suzuki, H., Yabuuchi, M., and Takahashi, S. (1996). A possible involvement of melanocortin 1-receptor in regulating feather color pigmentation in the chicken. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1308 (2)**: 164-168.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22 (22)**: 4673-4680.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14 (2)**: 178-192.

Turner, S.D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software* **3 (25)**: 731.

Våge, D.I., Fuglei, E., Snipstad, K., Beheirn, J., Landsem, V.M., and Klungland, H. (2005). Two cysteine substitutions in the *MC1R* generate the blue variant of the arctic fox (*Alopex lagopus*) and prevent expression of the white winter coat. *Peptides* **26 (10)**: 1814-1817.

Våge, D.I., Klungland, H., Lu, D., and Cone, R.D. (1999). Molecular and pharmacological characterization of dominant black coat color in sheep. *Mammalian Genome* **10 (1)**: 39-43.

Våge, D.I., Lu, D., Klungland, H., Lien, S., Adalsteinsson, S., and Cone, R.D. (1997). A non-epistatic interaction of *agouti* and *extension* in the fox, *Vulpes vulpes*. *Nature Genetics* **15**: 311-315.

van Dijk, E.L., Jaszczyszyn, Y., and Thermes, C. (2014). Library preparation methods for next-generation sequencing: tone down the bias. *Experimental Cell Research* **322 (1)**: 12-20.

van Soest, R., and van Bree, P. (1969). On the moult in the stoat, *Mustela erminea* Linnaeus, 1758, from the Netherlands. *Bijdragen tot de Dierkunde* **39 (1)**: 63-68.

van Zyll de Jong, C. (1992). A morphometric analysis of cranial variation of Holarctic weasels (*Mustela nivalis*). *Zeitschrift für Säugetierkunde* **57 (2)**: 77-93.

Varpe, Ø. (2017). Life history adaptations to seasonality. *Integrative and Comparative Biology* **57 (5)**: 943-960.

Vaurie, C. (1965). *The birds of the palearctic fauna: a systematic reference, non-passeriformes*. London: H.F. & G. Witherby.

Vrieling, H., Duhl, D.M., Millar, S.E., Miller, K.A., and Barsh, G.S. (1994). Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. *Proceedings of the National Academy of Sciences* **91 (12)**: 5667-5671.

Wandeler, P., Hoeck, P.E.A., and Keller, L.F. (2007). Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution* **22 (12)**: 634-642.

Westbury, M.V., Hartmann, S., Barlow, A., Wiesel, I., Leo, V., Welch, R., Parker, D.M., Sicks, F., Ludwig, A., Dalén, L.*, et al.* (2018). Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the brown hyena. *Molecular Biology and Evolution* **35 (5)**: 1225-1237.

Wilson, S.M., Yip, R., Swing, D.A., O'Sullivan, T.N., Zhang, Y., Novak, E.K., Swank, R.T., Russell, L.B., Copeland, N.G., and Jenkins, N.A. (2000). A mutation in Rab27a causes the vesicle transport defects observed in *ashen* mice. *Proceedings of the National Academy of Sciences* **97 (14)**: 7933-7938.

Wu, X., Rao, K., Bowers, M.B., Copeland, N.G., Jenkins, N.A., and Hammer, J.A. (2001). Rab27a enables myosin Va-dependent melanosome capture by recruiting the myosin to the organelle. *Journal of Cell Science* **114 (6)**: 1091-1100.

Young, N.D. (1996). QTL mapping and quantitative disease resistance in plants. *Annual Review of Phytopathology* **34 (1)**: 479-501.

Zhu, M., and Zhao, S. (2007). Candidate gene identification approach: progress and challenges. *International Journal of Biological Sciences* **3 (7)**: 420-427.

Zima, J., and Cenevova, E. (2002). Coat colour and chromosome variation in central European populations of the weasel (*Mustela nivalis*). *Folia Zoologica* **51 (4)**: 265-274.

Zimova, M., Hackländer, K., Good, J.M., Melo-Ferreira, J., Alves, P.C., and Mills, L.S. (2018). Function and underlying mechanisms of seasonal colour moulting in mammals and birds: what keeps them changing in a warming world? *Biological Reviews* **93 (3)**: 1478-1498.

Zimova, M., Mills, L.S., Lukacs, P.M., and Mitchell, M.S. (2014). Snowshoe hares display limited phenotypic plasticity to mismatch in seasonal camouflage. *Proceedings of the Royal Society B: Biological Sciences* **281 (1782)**: 20140029.

Zimova, M., Mills, L.S., and Nowak, J.J. (2016). High fitness costs of climate change-induced camouflage mismatch. *Ecology Letters* **19 (3)**: 299-307.

Appendices

Appendix I – Sampling information

Table S1 – Samples of *Mustela nivalis* from Sweden used for whole-genome sequencing. Sample codes refer to accession numbers of the Swedish Museum of Natural History (NRM), except when marked with asterisk (*). For each sample, information is given about colouration type, winter colour, percentage of body of white colour (% white), sex, and collection date and locality.

NRM Sample Code	Colour type	Winter colour	% white	Collection Date	Sex	Province	Locality
557916	*nivalis*	white	100	05-01-1964	Male	Jämtland	Hammerdal, Åsen
557922	*nivalis*	white	100	19-02-1964	Male	Jämtland	Lit, Brevåg
557940	*nivalis*	white	100	29-01- 1964	Male	Jämtland	Hammerdal
557950	*nivalis*	white	100	02-02-1964	Female	Jämtland	Hammerdal
557953	*nivalis*	white	100	07-02-1961	Female	Jämtland	Hammerdal
557964	*nivalis*	brown	0	20-01-1904	Female	Jämtland	Högrun
557967	*nivalis*	brown	0	20-01-1964	Female	Jämtland	Högrun
557970	*nivalis*	white	95	23-01-1961	Female	Jämtland	Häggenås, Norderåsen
557971	*nivalis*	white	95	12-01-1964	Female	Jämtland	Häggenås
557982	*nivalis*	white	95	23-01-1961	Male	Jämtland	Häggenås, Norderåsen
557983	*nivalis*	white	95	01-1964	Male	Jämtland	Högrun
557985	*nivalis*	white	95	12-01-1964	Male	Jämtland	Häggenås
557989	*nivalis*	white	100	20-02-1960	Female	Lycksele Lappmark	Gardvik, Björkås
557996	*nivalis*	white	95	03-02-1960	Male	Lycksele Lappmark	Gardvik
557998	*nivalis*	white	100	30-01-1961	Male	Lycksele Lappmark	Gardvik, Björkås
557999	*nivalis*	white	100	30-01-1961	Male	Lycksele Lappmark	Gardvik, Björkås
558018	*nivalis*	white	100	09-12-1962	Male	Södermanland	Öster-Malma
558020	*nivalis*	white	100	24-01-1961	Male	Dalarna	Ludvika, Ludvika
558025	*nivalis*	white	100	24-01-1961	Male	Dalarna	Ludvika, Ludvika
558026	*nivalis*	white	100	30-12-1959	Male	Värmland	Näsberget, Rösberget
558043	*nivalis*	brown	0	11-12-1961	Male	Södermanland	Öster-Malma
558050	*vulgaris*	brown	0	20-01-1964	Male	Östergötland	Linköping, Linköping
558115	*vulgaris*	brown	0	09-02-1964	Male	Östergötland	Linköping, Linköping
558117	*vulgaris*	brown	0	12-12-1962	Male	Östergötland	Linköping, Smedstad
558121	*vulgaris*	brown	0	09-02-1964	Male	Östergötland	Linköping, Linköping
558124	*vulgaris*	brown	0	06-01-1965	Male	Östergötland	Linköping, Linköping
558127	*vulgaris*	brown	0	05-02-1964	Female	Östergötland	Linköping, Linköping
558132	*vulgaris*	brown	0	09-02-1964	Female	Östergötland	Linköping, Linköping
558160	*vulgaris*	brown	0	12-12-1961	Male	Södermanland	Ludgo, Öster-Malma
558162	*vulgaris*	brown	0	01-02-1964	Male	Södermanland	Ludgo, Öster-Malma
558164	*vulgaris*	brown	0	01-02-1964	Male	Södermanland	Ludgo, Öster-Malma
558165	*vulgaris*	brown	0	10-12-1961	Male	Södermanland	Ludgo, Öster-Malma
558167	*vulgaris*	brown	0	17-12-1961	Male	Södermanland	Ludgo, Öster-Malma
558168	*vulgaris*	brown	0	16-12-1961	Male	Södermanland	Ludgo, Öster-Malma
558172	*vulgaris*	brown	0	07-01-1962	Female	Södermanland	Ålberga, Sörby
558178	*nivalis*	white	100	08-01-1962	Female	Södermanland	Mälby
558179	*nivalis*	white	100	29-01-1960	Male	Södermanland	Eskilstuna
558182	*vulgaris*	brown	0	14-02-1960	Male	Dalsland	Laxarby, Heden
558183	*vulgaris*	brown	0	12-12-1962	Male	Dalsland	Tydje, Västanå
JMF15*	*nivalis*	brown	5	17-02-1961	Male	Uppland	Farentuna

Table S2 – Additional samples of *Mustela nivalis* from Sweden used for genotyping experiments. Sample codes refer to accession numbers of the Swedish Museum of Natural History (NRM). For each sample, information is given about colouration type, winter colour, percentage of body of white colour (% white), sex, and collection date and locality.

NRM Sample Code	Colour type	Winter colour	% white	Collection Date	Sex	Province	Locality
557900	*nivalis*	white	100	23-12-1963	Male	Jämtland	Hammerdal, Åsen
557915	*nivalis*	white	100	09-01-1964	Male	Jämtland	Hammerdal
557918	*nivalis*	white	100	13-02-1964	Male	Jämtland	Högrun
557941	*nivalis*	white	100	01-1964	Male	Jämtland	Högrun
557943	*nivalis*	white	100	20-02-1964	Male	Jämtland	Högrun
557945	*nivalis*	white	100	15-12-1963	Male	Jämtland	Häggenås
557966	*nivalis*	brown	0	25-11-1963	Female	Jämtland	Hammerdal, Åsen
557968	*nivalis*	brown	0	04-11-1963	Female	Jämtland	Hammerdal
557973	*nivalis*	brown	20	15-12-1963	Female	Jämtland	Hammerdal, Åsen
557975	*nivalis*	white	80	01-1964	Male	Jämtland	Högrun
557977	*nivalis*	white	80	20-02-1964	Male	Jämtland	Högrun
557978	*nivalis*	brown	5	26-01-1964	Male	Jämtland	Häggenås
557979	*nivalis*	brown	5	26-11-1963	Male	Jämtland	Lit, Brevåg
557980	*nivalis*	brown	20	05-12-1963	Male	Jämtland	Hammerdal, Åsen
557984	*nivalis*	white	95	23-01-1961	Male	Jämtland	Häggenås, Norderåsen
557986	*nivalis*	white	95	15-12-1963	Male	Jämtland	Hammerdal
557987	*nivalis*	white	95	07-12-1963	Male	Jämtland	Hammerdal
558000	*nivalis*	white	100	25-01-1960	Male	Lycksele Lappmark	Gardvik, Björkås
558004	*nivalis*	white	100	28-12-1959	Male	Lycksele Lappmark	Gardvik, Björkås
558008	*nivalis*	white	100	15-01-1961	Male	Hälsingland	Färila, Töva
558015	*nivalis*	white	100	15-01-1962	Male	Härjedalen	Vemdalen, Vemhån
558016	*nivalis*	white	100	15-01-1963	Male	Södermanland	Äkers Styckebruk
558017	*nivalis*	white	100	06-03-1964	Male	Södermanland	Ludgo, Öster-Malma
558021	*nivalis*	brown	0	22-11-1964	Male	Södermanland	Ludgo, Öster-Malma
558037	*nivalis*	white	95	02-1963	Male	Uppland	Färentuna
558039	*nivalis*	white	80	28-11-1959	Female	Värmland	Fryksände, Stranna
558042	*nivalis*	brown	20	28-12-1959	Female	Lycksele Lappmark	Gardvik, Björkås
558048	*nivalis*	brown	0	26-11-1963	Male	Jämtland	Hammerdal, Åsen
558052	*vulgaris*	brown	0	12-12-1961	Male	Södermanland	Ludgo, Öster-Malma
558099	*vulgaris*	brown	0	21-12-1961	Male	Västergötland	Lilleskog
558120	*vulgaris*	brown	0	18-12-1983	-	Östergötland	Linköping, Linköping
558122	*vulgaris*	brown	0	09-02-1964	Male	Östergötland	Linköping, Linköping
558148	*vulgaris*	brown	0	04-02-1964	Female	Uppland	Uppsala, St. Björkby
558156	*vulgaris*	brown	0	11-1961	Male	Södermanland	Ludgo, Öster-Malma
558158	*vulgaris*	brown	0	27-12-1961	Male	Södermanland	Ludgo, Öster-Malma
558161	*vulgaris*	brown	0	17-11-1961	Male	Södermanland	Ludgo, Öster-Malma
558163	*vulgaris*	brown	0	11-02-1926	Male	Värmland	Lästringe, Hagby
558166	*vulgaris*	brown	0	04-12-1962	Male	Södermanland	Ludgo, Öster-Malma
558170	*vulgaris*	brown	0	13-11-1961	Male	Södermanland	Ludgo, Öster-Malma
558175	*nivalis*	white	100	16-11-1959	Female	Södermanland	Eskilstuna, Strängnäs
558180	*nivalis*	white	95	02-02-1962	Male	Södermanland	Eskilstuna, Strängnäs
558186	*vulgaris*	brown	0	14-11-1961	Male	Södermanland	Ludgo, Öster-Malma
558187	*nivalis*	white	100	07-01-1964	Female	Jämtland	Häggenås

Table S3 – Samples of *Mustela nivalis* from Poland used for genotyping experiments. Sample codes refer to the collection of the Mammal Research Institute, Polish Academy of Sciences. Winter samples are marked with an asterisk (*). For each sample, information is given about colouration type, winter colour, and collection date and locality.

Sample Code	Colour type	Winter colour	Collection Date	Sex	Locality	Place
KZ080818	*vulgaris*	unknown	18-08-2008	Male	S Poland	Nowa Słupia
KZ100415	*vulgaris*	unknown	15-04-2010	-	Central Poland	Tychów
KZ251	*nivalis*	unknown	16-07-1999	Male	Białowieża Forest	Reski
KZ257	*vulgaris*	unknown	22-09-2000	Male	Central Poland	Niepust
KZ267	*nivalis*	unknown	01-10-2001	Female	Białowieża Forest	
KZ279	*nivalis*	unknown	13-09-2002	Male	Białowieża Forest	Reski
KZ284	*nivalis*	unknown	22-09-2002	Female	NE Poland	Biebrza (Gugny)
KZ285	*nivalis*	unknown	22-09-2002	Male	NE Poland	Biebrza (Gugny)
KZ286	*vulgaris*	unknown	12-10-2002	Male	Białowieża Forest	Reski
KZ287	*vulgaris*	unknown	16-10-2002	Male	Central Poland	Sieraków
KZ288	*vulgaris*	unknown	22-10-2002	Male	Central Poland	Łąki Famułkowskie
KZ292	*nivalis*	unknown	07-06-2003	Male	Białowieża Forest	BNP (Dyrekcyjny)
KZ295	*nivalis*	unknown	14-06-2003	Female	Białowieża Forest	BNP
KZ298	*nivalis*	unknown	26-07-2003	Female	Białowieża Forest	Towarowa, Laki
KZ306	*vulgaris*	unknown	21-08-2003	Female	Central Poland	Famułki Królewskie
KZ308	*nivalis*	unknown	24-08-2003	Female	NE Poland	Biebrza (Barwik)
KZ314	*vulgaris*	unknown	23-09-2003	Male	Białowieża Forest	BNP
KZ316	*nivalis*	unknown	27-09-2003	Male	Białowieża Forest	BNP (Orłowski)
KZ318	*vulgaris*	unknown	30-09-2003	Male	Central Poland	Pożary
KZ320	*vulgaris*	unknown	08-10-2003	Male	Białowieża Forest	Reski
KZ321	*nivalis*	unknown	17-10-2003	Male	NE Poland	Biebrza (Barwik)
KZ322	*nivalis*	unknown	19-10-2003	Male	NE Poland	Biebrza (Barwik)
KZ324	*nivalis*	unknown	20-10-2003	Male	NE Poland	Biebrza (Barwik)
KZ326	*nivalis*	unknown	29-11-2003	Female	Białowieża Forest	Narewka, Podolany
KZ366	*vulgaris*	unknown	29-08-2004	Male	Białowieża Forest	BNP
KZ367	*vulgaris*	unknown	01-09-2004	Male	Białowieża Forest	Reski
KZ376	*vulgaris*	unknown	25-09-2004	Male	Białowieża Forest	BNP (Cegielnia)
KZ385	*vulgaris*	unknown	25-10-2004	Male	Białowieża Forest	BNP (Cegielnia)
KZ388	*vulgaris*	unknown	28-10-2004	Male	Białowieża Forest	BNP
KZ458	*nivalis*	unknown	05-10-2005	Female	Białowieża Forest	BNP
MR136183	*nivalis*	unknown	20-07-1964	Male	Białowieża Forest	Topiło
MRI103650	*nivalis*	unknown	27-08-1971	Female	E Poland	Sobibór
MRI155673	*nivalis*	unknown	27-07-1987	Male	Białowieża Forest	BNP
MRI161184	*nivalis*	unknown	16-05-1992	-	E Poland	Wólka Zamkowa
MRI19672	*nivalis*	unknown	19-07-1970	Male	NE Poland	Płaska (Jary Biele)
MRI27063	*vulgaris*	unknown	-	-	W Poland	Goleniów
MRI30257/1657	*nivalis*	unknown	29-07-1962	Female	NE Poland	Żytkiejmy
MRI30435/1835	*nivalis*	unknown	02-08-1962	Female	NE Poland	Żytkiejmy
MRI36395	*nivalis*	unknown	24-07-1964	Male	Białowieża Forest	Topiło
MRI37429/669	*vulgaris*	unknown	23-08-1963	Female	S Poland	Góry Świętokrzyskie
MRI41492	*vulgaris*	unknown	18-07-1964	Female	SW Poland	Kalnica (Bieszczady)
MRI45143	*vulgaris*	unknown	18-07-1965	Female	SW Poland	Wojcieszów Górny

Table S3 (continued)

Sample Code	Colour type	Winter colour	Collection Date	Sex	Locality	Place
MRI481/4726	*vulgaris*	unknown	23-07-1967	Male	W Poland	Gryfice
MRI48469	*vulgaris*	unknown	31-08-1965	Female	NW Poland	Górowo Iławeckie
MRI59832/10322	*vulgaris*	unknown	23-08-1966	Female	NW Poland	Darżlubie
MRI61813/16427	*vulgaris*	unknown	16-09-1966	Male	W Poland	Międzyzdroje
MRI69800/5856	*vulgaris*	unknown	19-08-1967	Male	W Poland	Miedzychód
MRI78813/22636	*vulgaris*	unknown	15-08-1968	Female	W Poland	Krosno Odrzańskie
MRI104349*	*nivalis*	white	18-02-1972	Male	Białowieża Forest	BNP
MRI104350*	*nivalis*	white	18-02-1972	Male	Białowieża Forest	BNP
MRI108522*	*nivalis*	white	21-12-1972	Male	Białowieża Forest	Zwierzyniec
MRI152404*	*nivalis*	white	20-03-1986	Female	NE Poland	Biebrza (Barwik)
MRI34457*	*nivalis*	white	26-12-1963	Female	Białowieża Forest	Hajnówka
MRI34532*	*nivalis*	white	12-01-1964	Female	Białowieża Forest	BNP
MRI42030*	*nivalis*	white	15-12-1964	Male	Białowieża Forest	BNP
MRI51145*	*nivalis*	white	15-02-1966	Male	Białowieża Forest	BNP
MRI51148*	*nivalis*	white	13-02-1966	Male	Białowieża Forest	BNP
MRI51152*	*nivalis*	white	17-02-1966	Female	Białowieża Forest	BNP
MRI51424*	*nivalis*	white	21-03-1966	Female	Białowieża Forest	Zwierzyniec
MRI66797*	*nivalis*	white	24-11-1967	Female	Białowieża Forest	Zwierzyniec
MRI72621*	*nivalis*	white	11-04-1968	Male	Białowieża Forest	BNP
MRI7337*	*nivalis*	white	22-12-1954	Female	Białowieża Forest	BNP
MRI87940*	*nivalis*	white	25-11-1969	Female	Białowieża Forest	BNP

S, SW, E, NE, NW, W – cardinal points
BNP – Białowieża National Park

Table S4 – Samples of *Mustela nivalis* used by Frank (1985) in his crossing experiments used for genotyping experiments. Sample codes refer to accession numbers of the Zoological Research Museum Alexander Koenig. For each sample, information is given about the specimen number attributed by Frank (1985), the generation of the specimen, respective father and mother, colouration type, and sex.

Sample Code	Specimen no.	Generation	Father	Mother	Colour type	Sex
MAM 2008-0009	45	F1	41	Flicka	*vulgaris*	Male
MAM 2008-0010	46	F1	41	Flicka	*vulgaris*	Male
MAM 2008-0011	52	F2	49	Flicka	*vulgaris*	Female
MAM 2008-0012	51	F2	49	Flicka	*vulgaris*	Male
MAM 2008-0013	57	F2	49	Flicka	*vulgaris*	Female
MAM 2008-0015	55	F2	49	Flicka	*vulgaris*	Male
MAM 2008-0016	58	F2	49	Flicka	*nivalis*	Female
MAM 2008-0017	68	F2	49	Flicka	*vulgaris*	Male
MAM 2008-0018	71	F2	49	Flicka	*nivalis*	Male
MAM 2008-0019	54	F2	49	Flicka	*nivalis*	Female
MAM 2008-0020	61	F2	49	Flicka	*nivalis*	Male
MAM 2008-0021	66	F2	49	Flicka	*nivalis*	Male
MAM 2008-0022	56	F2	49	Flicka	*nivalis*	Female
MAM 2008-0023	73	F2	49	Flicka	*nivalis*	Male
MAM 2008-0024	67	F2	49	Flicka	*nivalis*	Male
MAM 2008-0025	78	F2	49	Flicka	*vulgaris*	Male
MAM 2008-0026	79	F2	49	Flicka	*nivalis*	Male
MAM 2008-0027	81	F3	69	Flicka	*nivalis*	Male
MAM 2008-0028	83	F3	69	Flicka	*nivalis*	Male
MAM 2008-0029	72	F2	49	Flicka	*nivalis*	Female
MAM 2008-0030	70	F2	49	Flicka	*nivalis*	Female
MAM 2008-0031	84	F3	67	Flicka	*nivalis*	Female
MAM 2008-0032	91	F3	69	Flicka	*nivalis*	Male
MAM 2008-0033	93	F3	69	Flicka	*nivalis*	Female
MAM 2008-0034	82	F3	69	Flicka	*nivalis*	Female
MAM 2008-0036	48	F1	41	Flicka	*vulgaris*	Male
MAM 2008-0037	47	F1	41	Flicka	*vulgaris*	Male
MAM 2008-0038	50	F1	41	Flicka	*vulgaris*	Female
MAM 2008-0039	53	F2	49	Flicka	*vulgaris*	Male
MAM 2008-0040	60	F2	49	Flicka	*vulgaris*	Female
MAM 2008-0041	90	F3	69	Flicka	*nivalis*	Male
MAM 2008-0042	94	F3	69	Flicka	*nivalis*	Female
MAM 2008-0043	Flicka	P	-	-	*nivalis*	Female

Generation:
P – parental specimen
F1 – specimens resulting from parental crossing
F2 – specimens resulting from the backcross of F1 *vulgaris* males with the parental female
F3 – specimens resulting from crossings of *nivalis* F2 males with the parental female

Appendix II – Demographic models

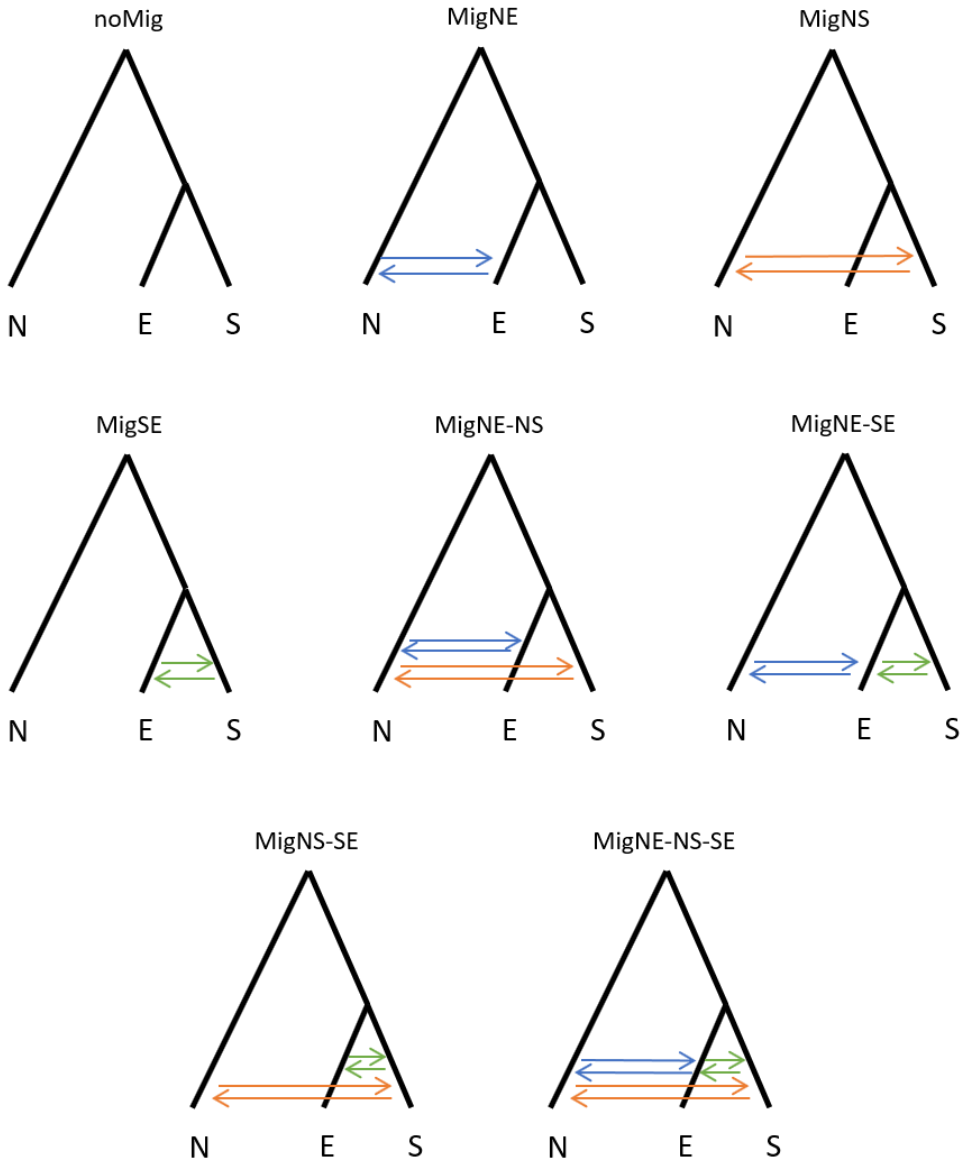

Figure S1 – Schematic diagram of all demographic models tested with fastsimcoal2, assuming as topology the population tree inferred with Treemix (N – Northern; E – Eastern; S – Southern). Effective population sizes were assumed constant through time and all migration events tested (indicated with arrows) were assumed asymmetric. Divergence times and effective population sizes were estimated for all models.

Appendix III – Detailed genotyping protocol

Genotyping was performed for 9 SNPs and a 12 bp deletion resorting to a total of 8 PCR reactions. Information regarding the variants amplified in each reaction and respective primers is shown in Table S5.

Table S5 – Variants targeted in each PCR reaction and primers used for amplification. Variants' positions are given in relation to scaffold 42 (GL896939.1) from the ferret reference genome. Forward primers are indicated through a negative distance to the variant, whereas reverse primers have a positive distance. Melting temperature (T_m) was calculated with NETPRIMER, using default parameters.

Amplicon	Variant(s) type	Variant(s) position	Primer name	Distance to variant(s) (bp)	Sequence (5' to 3')	T_m (ºC)
W1	SNP	118,068	sc42_118608F	-68	CCCACAGCTCACATACACTAAG	55.15
			sc42_118608R	19	CCAGGGCAGTCCTCGTC	55.92
W2	SNP	173,465	sc42_173465F	-55	TGCTGTTTCTCACTGTAGTTGCTC	59.27
			sc42_173465R	14	CCTAACCAAGGGGAGAGTGTG	58.66
W3	SNP + SNP	175,008 + 175,009	sc42_175008F	-53	CACGAGCACGTCAATGGC	58.39
			sc42_175008R	14	AGCAACATGCTGAAGATGGC	58.44
W4	deletion + SNP	182,432 – 443 + 182,449	sc42_182449F	-18	GACGTGGGGATACACAGGC	57.98
			sc42_182449R	60	CTTCGCAGTAAACCAGCACAG	58.70
W5	SNP	200,524	sc42_200524F	-55	TGACATCCCCATCAAGTAAGTG	57.44
			sc42_200524R	14	AATCCTGGCTCCACCTCTTC	58.01
W6	SNP	217,866	sc42_217866F	-55	GCTTGAGGCAATGAGGTAGATC	58.34
			sc42_217866R	14	GCTTGGATCTTACTGTCCTCCC	59.57
W7	SNP	275,086	sc42_275086F	-64	CTGTGAACAATTACGTCCCTTC	56.67
			sc42_275086R	12	GGGATCATTCACCACCTCC	56.64
W8	SNP	313,924	sc42_313924F	-51	GACTAGCTCTCAGGGAAAAGG	55.31
			sc42_313924R	13	TTGAAGCCCCACTAGTTGC	55.96

For each PCR reaction, reaction mixture and conditions were applied as follows (amplicons and primers as indicated in Table S5):

a) *Amplicon W1*: Reaction mixture for amplification of this fragment included 2.5 µL of Qiagen PCR MasterMix, 0.25 µL of each primer sc42_118608F and sc42_118608R at a concentration of 10 µM, and 2 µL of sample DNA at 2.5 ng/µL, in a total volume of 5 µL. PCR conditions consisted of an initial denaturation at 95ºC for 15 min, followed by 35 cycles of denaturation for 30 s at 95ºC, annealing for 40 s at 59ºC and extension for 30 s at 72ºC, followed by a final extension at 72ºC for 10 min. PCR product was expected with 125 bp.

b) Amplicon W2: Amplification reaction mixture used for this amplicon was the same used for amplicon W1, with exception to the primers used, sc42_173465F and sc42_173465R. PCR conditions were similar to W1 but with an annealing temperature of 63ºC. PCR product was expected with 113 bp.

c) Amplicon W3: Reaction mixture and PCR conditions applied to this amplicon were the same as for amplicon W2, with exception to the primers used, sc42_175008F and sc42_175008R. PCR product was expected with 108 bp.

d) Amplicon W4: Amplification reaction mixture included 1.7 µL of Qiagen PCR MasterMix, 0.10 µL of each primer sc42_182449F and sc42_182449R at 10 µM, 1.1 µL of deionised water, and 2 µL of sample DNA at 2.5 ng/µL, in a total volume of 5 µL. PCR conditions for amplification of the fragment consisted of an initial denaturation of 15 min at 95ºC, followed by a touchdown program of 10 cycles of denaturation at 95ºC for 30 s, annealing at 68ºC to 63ºC (with a 0.5ºC decrease in each cycle) for 5 s, and extension at 72ºC for 20 s, followed by 35 cycles of denaturation at 95ºC for 30 s, annealing at 63ºC for 30 s, and extension at 72ºC for 20 s, followed by a final extension at 72ºC for 10 min. PCR products were expected with 134 bp or 122 bp, according to the absence or presence of the deletion, respectively.

e) Amplicon W5: Reaction mixture for this fragment was the same applied for amplicon W1, except for the primers used, which were sc42_200524F and sc42_200524R. PCR conditions were similar to W1 but with an annealing temperature of 61ºC. PCR product was expected with 110 bp.

f) Amplicon W6: Amplification reaction mixture was composed by 2.5 µL of Qiagen PCR MasterMix, 0.10 µL of each primer sc42_217866F and sc42_217866R at 10 µM, 0.3 µL of deionised water, and 2 µL of sample DNA at 2.5 ng/µL, in a total volume of 5 µL. PCR conditions were identical to those of amplicon W1, with exception to the annealing temperature, which was of 68ºC. PCR product was expected with 109 bp.

g) Amplicon W7: For this amplicon, reaction mixture and PCR conditions were the same applied for amplicon W2, with exception for the primers used, which were sc42_275086F and sc42_275086R. PCR product was expected with 116 bp.

h) Amplicon W8: Amplification reaction mixture for this fragment was the same applied for amplicon W1, with exception for the primers used, which were sc42_313924F and sc42_313924R. PCR conditions were identical to the ones described for W1 but with an annealing temperature of 66ºC. PCR product was expected with 103 bp.

Resulting PCR products were analysed through an electrophoresis in a 2% agarose gel stained with 17.5 µL/L Gel Red and immersed in TBE (Tris-Borate-EDTA Buffer; Tris 89 mM, Boric Acid 89 mM and EDTA 2 mM, pH 0.8) buffer with a concentration of 0.5X. The gel was loaded with 1 µL of each PCR product combined with 3 µL of bromophenol blue. In each electrophoresis run, 2.5 µL of Marker V were also used. Runs were performed at 300 V and gels were visualised under ultraviolet light. Only samples with a visible band in the gel were used for sequencing.

Selected PCR products were cleaned using two enzymes, Exonuclease I (Exo) and Shrimp Alkaline Phosphatase (SAP). Generally, for each 2 µL of PCR product, 0.6 µL of ExoSAP (mixture previously prepared with 1 µL of Exo to each 2 µL of SAP) were used. Samples were incubated in a thermocycler at 37ºC for 15 min, followed by 15 min at 80ºC to inactivate enzymes.

Cleaned products were submitted to a cycle sequencing reaction. Reaction mixture included 0.25 µL of BigDye™ Terminator 3.1 Ready Reaction Mix, 0.5 µL of BigDye™ Terminator 5X Sequencing Buffer (Applied Biosystems), 0.27 µL of the adequate primer (i.e., for each amplicon, the primer that had the higher absolute distance to variant(s) – see Table S5) 10 µM concentrated, 3.48 µL of deionised water, and 0.5 – 1 µL of DNA (cleaned PCR product). The DNA volume used for each sample was determined according to the intensity of the band visualised in the electrophoresis gel. Conditions for the cycle sequencing reaction were as follows: an initial denaturation step at 96ºC for 3 min, followed by 25 cycles of denaturation at 96ºC for 10 s, annealing at 58ºC for 5 s and extension at 60ºC for 4 min.

Resulting products were cleaned with a Sephadex protocol. Columns were filled with 200 µL of Sephadex (66.7 g/L) and centrifuged at 2,950 rotations per minute (rpm) for 3 min to obtain a resin matrix to filter DNA. Then, DNA was added to the columns and centrifuged under the same conditions. Final cleaned sequences were dried at 95ºC for 30 min and rehydrated with 12.5 µL of formamide.

Appendix IV – Sequencing statistics per sample

Table S6 – Summary of sequencing statistics per individual. Number of reads is given for both sequencing output and trimmed data. Percentage of properly mapped reads is given for both the ferret reference and the pseudoreference for trimmed data, and for the pseudoreference after removing duplicates. Percentage of duplicates (Dup) and final coverage (Cov) are also shown.

Sample	Number of reads		Mapping ferret (%)	Mapping pseudoref. (%)		Dup (%)	Cov (X)
	Raw data	Trimmed data	Trimmed data	Trimmed data	Data without duplicates		
557916	48,286,147	48,072,976	85.98	87.21	85.25	16.25	0.94
557922	46,224,849	45,853,951	79.00	80.68	76.84	21.05	0.65
557940	54,375,671	54,118,741	84.15	85.49	82.85	18.34	1.01
557950	46,319,834	46,125,071	86.20	87.44	85.48	16.45	0.91
557953	54,601,887	53,905,636	77.95	79.91	76.83	16.79	0.78
557964	53,158,846	52,893,036	85.74	87.01	84.80	17.20	1.02
557967	45,112,983	44,971,368	88.04	89.14	87.40	16.61	0.95
557970	47,787,617	47,610,768	88.00	89.05	87.29	16.54	0.99
557971	46,723,441	46,500,874	85.98	87.21	85.23	15.90	0.94
557982	48,464,566	48,314,369	87.07	88.21	86.25	16.93	1.02
557983	57,661,697	57,435,894	86.43	87.62	85.52	17.28	1.17
557985	43,211,612	43,019,389	85.33	86.66	84.63	16.03	0.82
557989	51,534,895	51,261,087	82.51	83.97	80.38	22.30	0.81
557996	48,600,665	48,336,499	84.09	85.47	83.02	17.78	0.87
557998	48,958,292	48,749,815	85.25	86.57	84.11	17.99	0.93
557999	53,238,966	53,068,082	87.43	88.48	86.41	17.47	1.12
558018	62,478,645	61,666,982	78.32	80.07	76.82	18.24	0.94
558020	46,525,977	46,329,720	86.27	87.47	85.48	16.78	0.93
558025	47,862,417	47,642,981	86.75	87.89	85.58	19.23	1.02
558026	49,132,190	48,822,701	83.31	84.68	82.01	18.33	0.85
558043	61,044,928	60,772,502	85.05	86.35	83.92	18.33	1.10
558050	52,278,249	52,083,418	87.54	88.63	86.14	21.46	1.12
558115	53,672,605	53,483,607	85.36	86.49	83.35	22.81	0.99
558117	46,838,412	46,626,967	86.76	87.85	85.31	20.00	0.98
558121	44,692,725	44,558,697	88.23	89.19	86.90	19.88	0.98
558124	48,837,308	48,700,238	88.36	89.31	87.49	17.41	1.06
558127	48,848,812	48,583,515	85.17	86.53	83.79	20.05	0.94
558132	46,786,217	46,569,606	86.24	87.45	84.64	21.18	0.93
558160	48,856,782	48,664,634	87.44	88.55	86.30	19.27	1.05
558162	43,228,179	43,037,993	86.60	87.71	85.52	18.10	0.88
558164	54,522,855	54,168,476	82.49	83.96	81.32	17.89	0.95
558165	53,618,898	53,344,251	85.74	87.01	85.01	16.20	1.08
558167	56,450,021	56,294,370	89.32	90.23	88.68	15.65	1.33
558168	48,454,822	48,215,406	82.88	84.19	81.25	19.49	0.87
558172	46,460,906	46,202,431	85.64	86.89	84.55	18.43	0.91
558178	44,445,660	44,214,449	84.93	86.31	84.20	16.11	0.84
558179	54,153,980	53,961,292	86.66	87.85	85.73	18.03	1.05
558182	56,219,223	55,989,858	85.59	86.77	84.53	17.69	1.10
558183	55,837,632	55,544,861	85.33	86.63	83.91	20.49	1.07
JMF15	49,458,742	49,251,737	87.54	88.67	86.59	18.65	1.10
MEAN	50,374,204	50,124,206	85.42	86.67	84.28	18.27	0.98

Appendix V – NGSadmix analysis

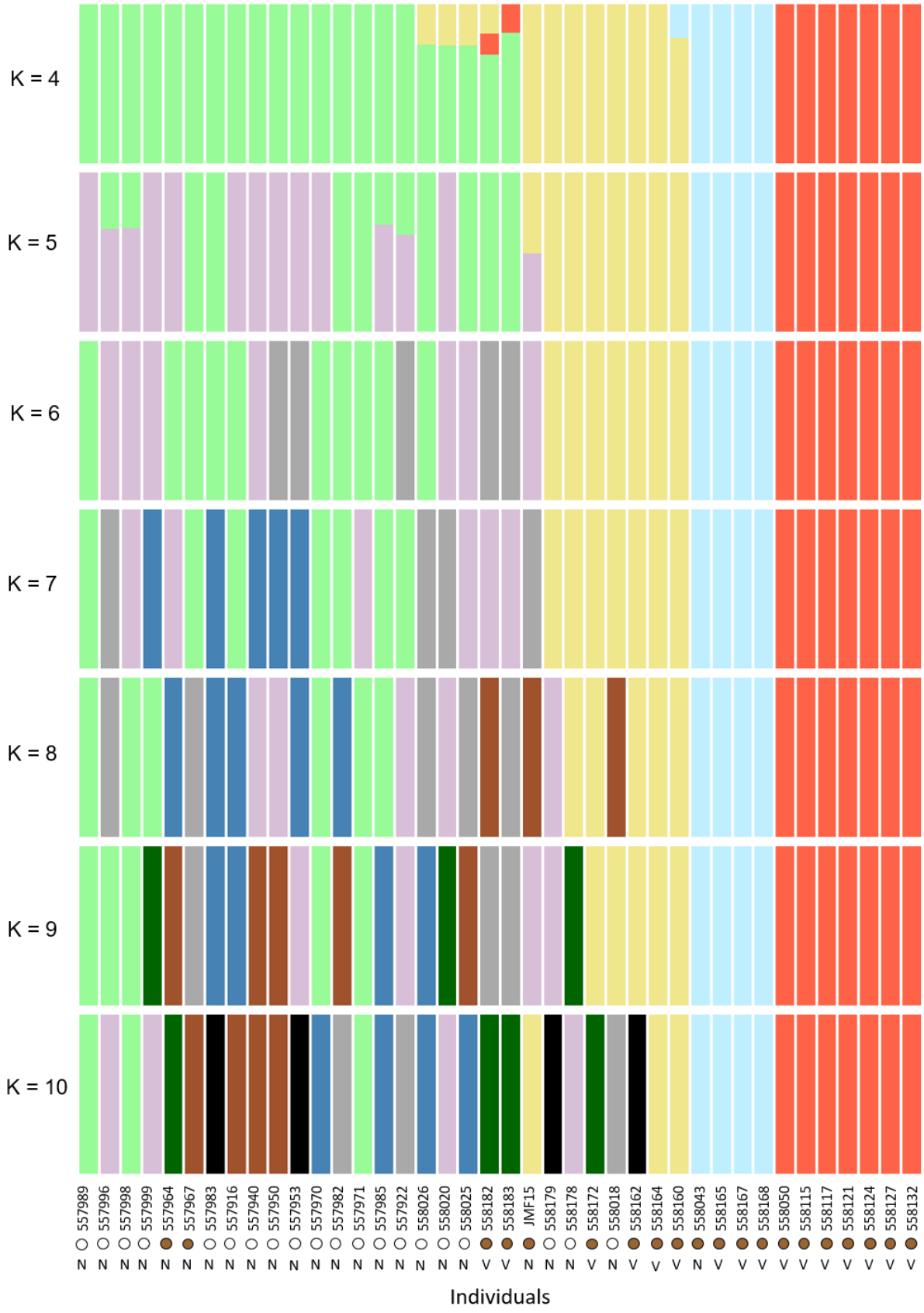

Figure S2 – Admixture proportions of the population of *Mustela nivalis* from Sweden, inferred with NGSadmix, for K ranging between 4 and 10. The best likelihood run is depicted for each K but no convergence between the highest likelihood runs was achieved within each value. The winter colour (brown or white) and colouration type (*nivalis* – N or *vulgaris* – V) of each specimen are also depicted.

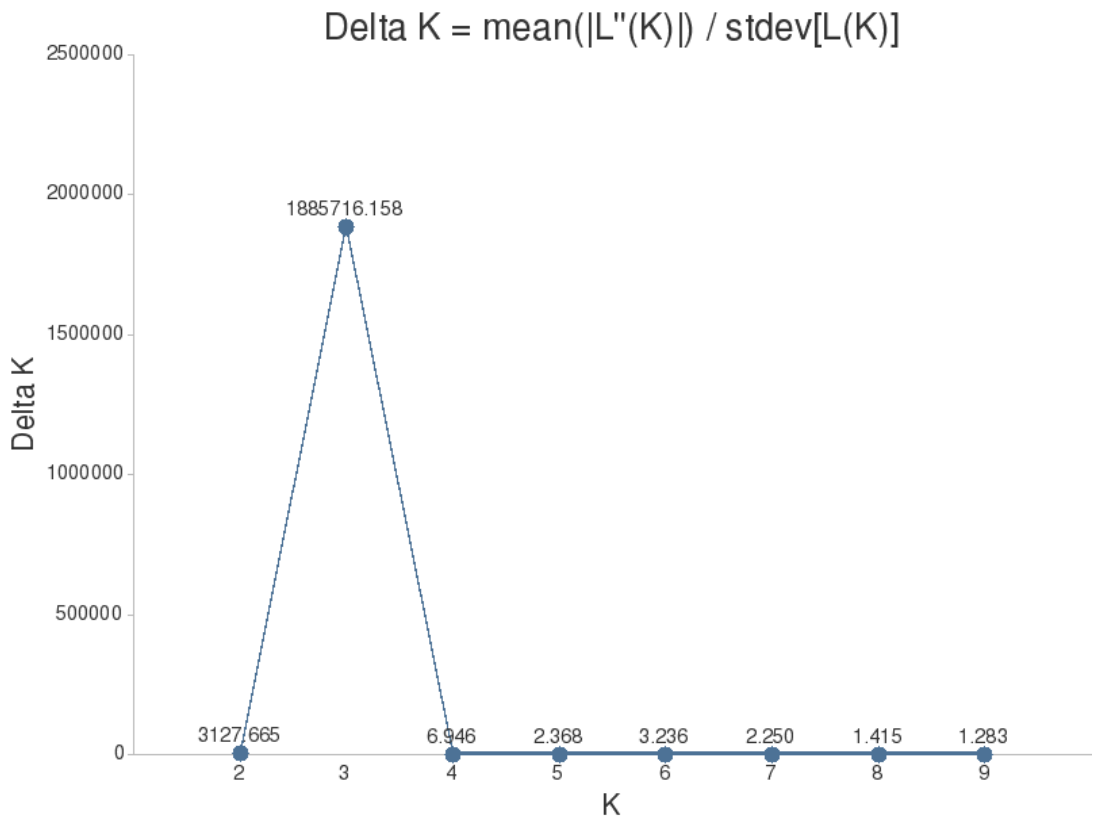

Figure S3 – Best K value inferred based on the results of NGSadmix analysis, using Clumpak. For each K value between 2 and 9, the value of Delta K (ΔK, as defined in Evanno *et al.*, 2005) is shown. K = 3 was inferred as the best number of ancestral populations.

Appendix VI – Demographic inferences

Table S7 – Selection of the best demographic model through the Akaike Information Criterion (AIC). For each model, the maximum log_{10} likelihood based on 50 independent optimisations, the number of estimated parameters (d), the computed AIC value, the difference to the lowest AIC (Δ), and the normalised relative likelihood (*w*) are shown. Models are named according to Figure S1 – Appendix II.

Model	log10(Lhood)	d	AIC	*Δ*	*w*
MigNE-SE	-415,408.93	11	1,913,050.819	0	1
MigNS	-415,418.68	9	1,913,091.720	41	1.31E-09
noMig	-415,430.578	7	1,913,142.512	91.693	1.23E-20
MigSE	-415,436.527	9	1,913,173.908	123.09	1.87E-27
MigNE	-415,440.643	9	1,913,192.863	142.04	1.43E-31
MigNE-NS	-415,441.047	11	1,913,198.724	147.90	7.64E-33
MigNE-NS-SE	-415,451.266	13	1,913,249.784	198.96	6.24E-44
MigNS-SE	-415,470.573	11	1,913,334.696	283.88	2.28E-62

N – Northern, S – Southern, E – Eastern

Table S8 – Parameters point estimates for the best fitted demographic scenario used for estimation of optimised jSFS. Point estimates are those of the maximum likelihood run for the best model. Times of divergence are given in generations, and migration rates are given in the migration fraction per generation.

Parameter	Point estimation
nPOP_N	3,602
nPOP_S	1,294
nPOP_E	604
nANC_SE	10,021
nANC	1,162,887
TDIV_SE	89
TDIV_N	118
MIG_EN	4.77E-04
MIG_NE	1.79E-03
MIG_ES	7.15E-04
MIG_SE	5.80E-03

nPOP – effective population size (EPS) of each population (N – Northern, S – Southern, E – Eastern); nANC_SE – EPS of the ancestral between S and E; nANC – EPS of the ancestral population; TDIV_SE /_N – time of divergence between S and E or S/E and N; MIG_EN /_NE – migration rate from E to N or N to E; MIG_ES /_SE – migration rate from E to S or S to E.

Appendix VII – Genes identified from whole-genome scans

Table S9 – Genes annotated within or near windows included in the top 0.01% highest F_{ST} values, averaged in 10 kb non-overlapping windows, for both winter-brown vs. winter-white (WB) and *nivalis* vs. *vulgaris* (NV) estimates. Windows are identified by their middle position. For each window, the corresponding scaffold and F_{ST} estimates are indicated. Candidate regions, as defined in the main text, are highlighted in bold.

Scaffold no.	Scaffold name	Window	F_{ST} - NV	F_{ST} - WB	Annotated genes
1	GL896898.1	39245000		0.36022	FYB1
13	GL896910.1	19585000		0.40495	TANGO6
19	GL896916.1	18075000		0.36349	TBC1D20
22	GL896919.1	1025000		0.35992	*
42	**GL896939.1**	**175000**	0.42787		MC1R TCF25 SPIRE2 FANCA ZNF276
		185000	0.43560		
		205000	0.51336	0.43923	
		215000	0.41114		
		225000	0.40813		
		255000	0.42597		
		265000	0.48731	0.40123	
		275000	0.43080		
		295000	0.41207		
		305000		0.38608	
		315000	0.54257	0.42207	
		325000	0.41951	0.35498	
50	GL896947.1	245000		0.40570	HLTF
68	GL896965.1	905000	0.39243		*
		4845000	0.43171		*
		6545000	0.39226	0.36922	REPS2
76	GL896973.1	6835000	0.46193	0.39690	NEURL4 upstream ACAP1[a]
84	GL896981.1	7665000		0.37037	*
86	GL896983.1	5885000		0.38710	IL13RA1
		5905000		0.40256	
108	GL897005.1	5765000		0.46832	CWC27
121	GL897018.1	5905000	0.41663		downstream CABS1[a] predicted protein
122	GL897019.1	1905000		0.36799	SEC14L4
128	GL897025.1	3525000	0.38972		upstream ERBB4[a]
166	GL897063.1	2415000	0.39815		*
206	**GL897103.1**	**2225000**	0.43439		*
		2235000	0.39592		
234	**GL897131.1**	**1395000**		0.35131	DRG2
		1405000	0.44256	0.44007	
		1415000	0.40249		
236	GL897133.1	895000		0.35214	upstream GPR50[a]
485	GL897371.1	85000		0.35133	*
490	GL898387.1	5000		0.37882	*

* No genes were annotated within or near these windows
[a] Gene located less than 100 kb out of the limits of the window

Appendix VIII – SNPs along the strongest candidate region

Table S10 – SNPs identified in the candidate region of scaffold 42 (GL896939.1) plus 100 kb flanking regions that follow the expected inheritance pattern of winter morphs (following Frank, 1985). For each SNP, position in the scaffold, the allele fixed in white individuals, and the alternative allele are shown. Estimated frequencies for the white allele (in winter-white individuals) and for the alternative allele (in winter-brown specimens) and the number of specimens represented in each group are also presented.

Position	Allele		Winter-white specimens		Winter-brown specimens	
	white	alternative	Freq. white	no. Ind	Freq. alternative	no. Ind
118608	G	A	0.999995	11	0.792330	13
121119	G	A	0.999991	12	0.603033	10
147910	A	C	0.999995	8	0.910957	10
155694	C	T	0.999992	11	0.486132	10
161685	G	A	0.999992	9	0.400142	9
168107	C	G	0.999995	10	0.681168	9
173465	G	A	0.999992	9	0.602248	12
174080	G	A	0.999993	16	0.878097	7
175008	A	T	0.999995	9	0.699241	13
175009	G	T	0.999995	9	0.699241	13
182347	A	G	0.999997	9	0.879006	13
182449	A	C	0.999994	11	0.606014	10
192089	T	C	0.999995	10	0.479673	12
195592	G	C	0.999997	7	0.696336	10
196023	G	T	0.999996	12	0.727235	14
200524	A	C	0.999997	9	0.788259	13
204832	T	G	0.999997	8	0.593443	12
206601	G	A	0.999996	7	0.628771	8
208809	G	A	0.999992	8	0.686785	12
209427	T	C	0.999992	13	0.612379	10
210963	T	A	0.999998	8	0.646301	10
213804	C	T	0.999995	9	0.862934	12
215086	C	A	0.999996	12	0.713674	11
217866	G	A	0.999996	13	0.597195	14
220024	C	A	0.999996	8	0.669280	8
227509	G	A	0.999994	12	0.875333	7
237920	G	A	0.999997	8	0.604868	13
244782	G	A	0.999996	8	0.845306	6
253063	C	T	0.999994	10	0.766569	12
253509	C	T	0.999995	10	0.746513	12
254822	G	A	0.999993	8	0.647957	10
255535	C	T	0.999997	12	0.676275	10
263011	C	A	0.999992	6	0.811504	13
263419	C	T	0.999997	8	0.715580	11
267870	G	C	0.999992	9	0.902175	9
268625	T	G	0.999997	12	0.763108	12

Table S10 (continued)

Position	Allele		Winter-white specimens		Winter-brown specimens	
	white	alternative	Freq. white	no. Ind	Freq. alternative	no. Ind
273824	G	A	0.999996	7	0.495871	13
275086	G	A	0.999997	11	0.674998	14
277287	T	A	0.999996	14	0.751052	15
278406	T	C	0.999994	11	0.802214	13
282409	T	G	0.999995	7	0.698137	12
287119	A	T	0.999996	8	0.732651	12
290656	C	T	0.999994	7	0.877319	8
291527	T	C	0.999995	7	0.743438	10
296231	G	C	0.999997	10	0.798278	13
296371	C	T	0.999993	8	0.802243	12
299168	C	T	0.999992	7	0.738295	12
303999	T	C	0.999995	11	0.753707	12
304520	G	A	0.999994	12	0.752070	10
306067	A	G	0.999997	8	0.663445	10
306158	G	C	0.999996	10	0.745480	7
310055	C	G	0.999997	8	0.736270	16
310358	A	C	0.999995	9	0.743365	10
310362	T	G	0.999996	10	0.671917	11
310626	T	C	0.999994	12	0.869495	12
311514	G	A	0.999994	11	0.605351	11
313127	A	C	0.999992	8	0.927463	12
313924	T	G	0.999997	9	0.652668	12
315679	T	C	0.999996	6	0.844336	10
320895	C	T	0.999998	8	0.572409	10
321146	A	G	0.999998	11	0.762954	8
342677	A	G	0.999997	11	0.535300	13
348271	G	A	0.999995	7	0.625042	8
353558	G	T	0.999997	16	0.686756	13
366083	G	A	0.999997	13	0.571159	9
373121	C	A	0.999995	8	0.509133	10
374155	G	A	0.999996	6	0.881031	8
381698	G	A	0.999994	11	0.418099	11
389398	C	T	0.999993	10	0.904703	10
391483	G	C	0.999991	6	0.662254	12
396696	G	A	0.999996	10	0.653703	6
407051	C	T	0.999995	9	0.626031	12
407388	C	T	0.999992	7	0.712791	10
410395	G	A	0.999992	9	0.452988	12
418260	G	A	0.999995	7	0.461607	8
420438	A	T	0.999993	8	0.416020	6
421494	G	A	0.999996	8	0.527186	6
421994	G	A	0.999994	7	0.628768	9
425112	C	T	0.999996	7	0.674762	12
428725	T	G	0.999997	6	0.513773	11

Appendix IX – Selection tests for the identified genetic populations

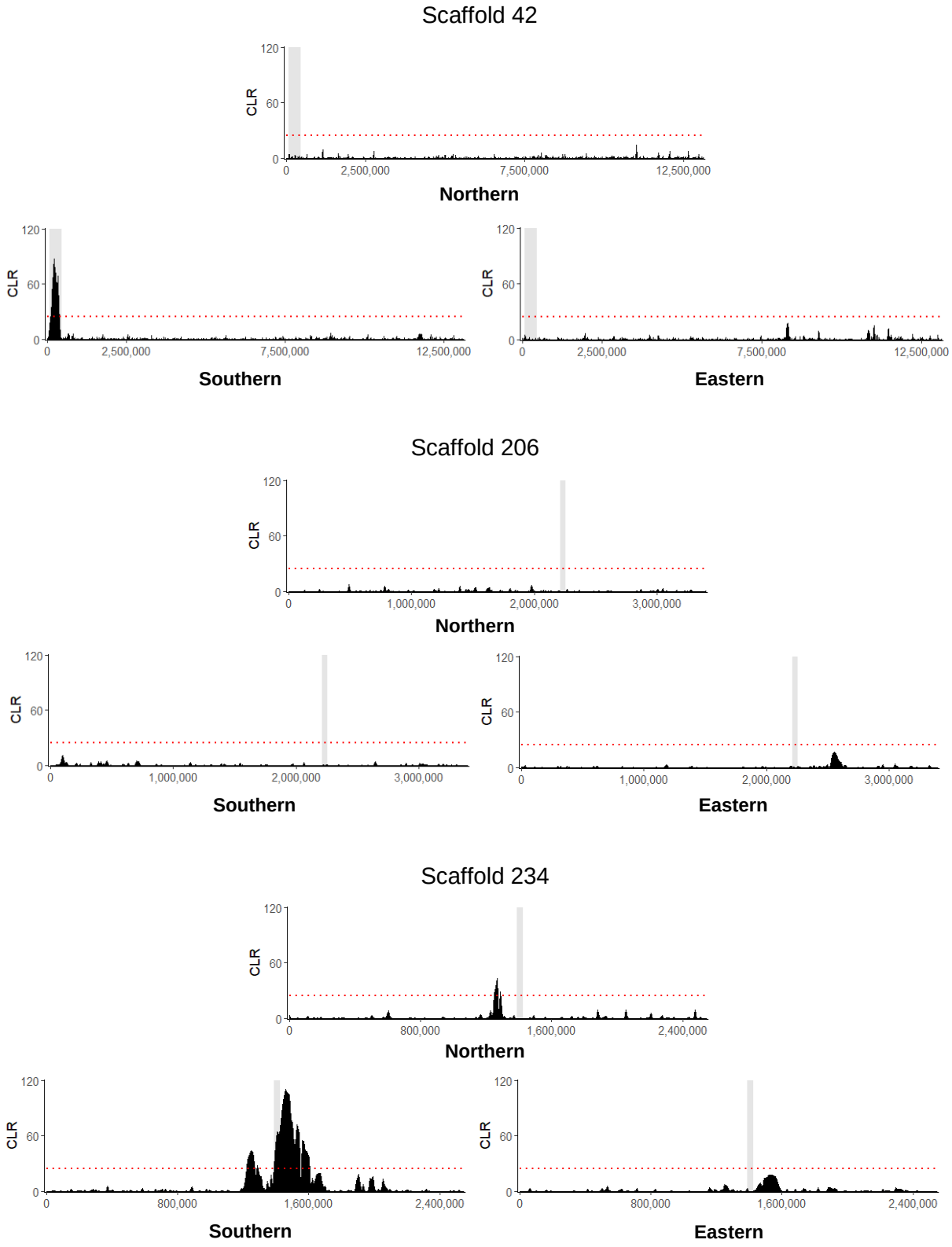

Figure S4 – Inferred selective sweeps across the three candidate regions at scaffold 42 (GL896939.1), scaffold 206 (GL897103.1) and scaffold 234 (GL897131.1) for the three genetic populations identified in *Mustela nivalis* from Sweden. Grey shades represent the identified candidate regions. Red lines represent the 99th percentile of estimated CLR values