

Nina Casillas

NTNU
Norwegian University of
Science and Technology
Faculty of Natural Sciences
Department of Biology

Nina Casillas

Phylogeography of the European adder

Investigating a rare phenotype in an isolated island
population

June 2022



Norwegian University of
Science and Technology

Phylogeography of the European adder

Investigating a rare phenotype in an isolated island population

Nina Casillas

Nordic Master in Biodiversity and Systematics

Submission date: June 2022

Supervisor: Dr. Mike Martin

Co-supervisor: Dr. José Cerca

Norwegian University of Science and Technology
Department of Biology

Abstract

Color polymorphism and the genetic basis driving pattern variation in ectothermic squamates is greatly under-studied outside of melanocytes. Variation in color and pattern evolution in wild populations of squamates is a highly informative feature that reflects details about defense strategies, feeding habits, environmental factors such as temperature, as well as overall fitness. *Vipera berus* is the most widely distributed and northernmost terrestrial snake in the world, yet they show very little variation in morphology across their range. In this study, we investigated the genetic basis of a rare phenotype not observed to this degree in the entire range of *V. berus* using a combination of whole-genome shotgun sequencing, pairwise F_{ST} estimates, genome-scanning and a genome-wide association based on 3,530,627 SNPs and 154 haplotypes. Seventeen populations of *Vipera berus*, as well as five outgroup *Vipera* species, were sampled within twelve countries across Eurasia to analyze the population structure of European vipers. Using these methods, we discovered 15 candidate regions, five candidate genes, four potential candidate genes and six SNPs significantly associated with the striped pattern in the isolated island population of Gossa. The most notable of these was the discovery of PMEL, a well known key component of melanosome pigmentation. Additionally, we compared two subspecies of *Vipera berus* to investigate why the nominate subspecies had greater success in colonizing their vast range. This resulted in the discovery of 24 candidate regions, five candidate genes and nine potential candidate genes that significantly differentiated the two subspecies. Of these, SVMP was identified as a potential candidate gene, which has been suggested to improve digestion efficiency of snakes in colder climates. Limitations of this experiment were primarily due to a lack of access to a thorough genome annotation for *Vipera berus*, which prevented the identification of >70% of candidate regions and five significant SNPs using the genome annotation. Although these genes were not identified using the annotation, they broaden the possibility of targeting these regions in the future and contribute to our understanding of genetic variation as well as pigmentation genes among isolated wild populations of snakes. Our results provide insight into the parallel evolution and genetic basis of pigmentation between ectothermic squamates using the unbiased approach of genome scanning, a method that has rarely been used in this context.

Acknowledgements

I have gained an invaluable amount of knowledge and experience from the hard-working, exceptionally friendly, and intelligent people in the NTNU Holomuseomics group. I can not express my gratitude enough for how welcoming and informative everyone has been throughout my time here. I would first like to acknowledge my wonderful supervisors: Dr. Mike Martin and Dr. José Cerca who were endlessly informative and encouraging from giving me the opportunity to catch snakes out in the field, to long hours of bioinformatic analyses.

I would like to thank Kjeld Henrik Ophus who guided me through the collection of over 70 snakes, sharing his infectious passion for *Vipera berus* at every moment possible. Thank you to Dr. Vanessa Bieker, Dr. Jaelle Brealey, Dr. Sarah Martin, and Jaime Lagos for their contributions both in the lab and digitally. In addition I would like to thank Dr. Dag Dolmen, Dr. Dirk Bauwens (Institute of Nature and Forest Research), Dr. Sarah Ball, Dr. Thomas Madsen (Deakin University), Dr. Sylvain Ursenbacher (University of Basel), and Dr. Oleksandr Zinenko (V. N. Karazin Kharkiv National University) who were kind enough to share their knowledge and collections with me.

I would also like to thank my family, my sweet dog Logan and my loving partner Mark, who continues to push me to dream bigger and do more than I could have ever imagined alone. Finally, I would like to acknowledge my beautiful Tía Mona who unfortunately passed away recently.

Contents

Abstract	1
Acknowledgements	2
Introduction	4
1.1 Vipera berus	5
1.2 Phylogeography of Vipera berus	5
1.3 Vipera berus taxonomy	6
1.4 Interesting morphology on Gossa	7
1.5 Vipera aspis and the pro-opiomelanocortin (POMC) gene	7
1.6 Pigmentation and pattern formation	8
1.7 Experimental design	9
Methods	10
2.1. Sample acquisition	10
Sampling on Gossa	11
2.2 DNA extraction, library preparation and sequencing	11
2.3 Bioinformatics analysis	11
2.4 Population structure and phylogenetics	12
2.5 FST Analysis	13
2.6 Genome-wide association study (GWAS)	14
Results	15
3.1 Population structure and phylogeny	16
3.2 FST Analyses	20
Striped vs. unstriped	20
Eurasian populations	23
V. b. berus vs. V. b. bosniensis	25
3.3 Genome-wide association study (GWAS)	28
Discussion	29
References	35
Supplementary Information	43
SI 1: DNA extraction	43
SI 2: Library build protocol	43
SI 3: Methods and models used in the methods and materials: will fill this in later**	45
Supplementary figures	46
SF 1: Isohelix buccal swab instructions	46

Introduction

The mechanisms of intraspecific melanin-based polymorphism has been a mesmerizing topic for researchers of diverse taxa seeking to answer questions regarding sexual selection, predator-prey interaction, thermoregulation, ecological niches, pattern formation, mimicry, etc., while being easily distinguishable and therefore tractable (Roulin 2004; Roulin 2014; MURRAY 2011; Murray & Myerscough 1991; Mckinnon & Pierotti 2010). Pattern and color variation of reptiles has especially been of interest because of the wide range of diversity which are not present in other animals (Murray 2011; Murray & Myerscough 1991). As color polymorphism often indicates there are additional changes in traits differentiating morphologies, (Mckinnon & Pierotti 2010), many studies find associations between intraspecific melanin-based polymorphism and alternative behavioral or mating strategies (Martínez-Freiría et al. 2020).

With the genetic mechanisms of the diversity of color polymorphism being heavily studied in domestic mammals and birds (Krude et al. 1998; Sutton et al. 2005; Bissig et al. 2016; Hellström et al. 2011; Theron et al. 2001), the research has a heavy focus on melanocytes, their only color-producing cell. Since snakes also have melanocytes, there have been various studies on genes associated with melanocytes in snakes (Ducrest et al. 2014; Cox et al. 2013; Corso et al. 2012; Rosenblum et al. 2004). Although melanocytes do contribute to pattern formation, reptiles have additional pigmentation cells within their dermis called chromatophores, which also contribute to color and pattern (Ullate-Agote et al. 2020; Alibardi 2013). As most pigmentation and pattern formation has been conducted with melanocytes in mind, little research on reptile polymorphism targets genes relative to chromatophores.

With the European adder, *V. berus*, showing very little variation in color or pattern morphology (i.e. melanistic and concolor; Andren & Nilson 1981; Cui et al. 2016) across its impressive range (Ursenbacher et al. 2006), the recent discovery of a rare striped pattern in an isolated island population off the coast of Norway has raised questions on the importance of their pattern morphology. Other European vipers such as *V. aspis* and *V. seoanei* show relatively high levels of polymorphism throughout their range, exhibiting melanistic, concolor, striped and zig-zag patterns. With *V. berus* occupying the northernmost range of any terrestrial snake, why have they not opted for a higher proportion of melanism like their southern relatives? Several studies have shown melanistic morphs to experience higher predation, but more efficient thermoregulation (Martínez-Freiría et al. 2017). What has prevented *V. berus* from developing similar proportions of color polymorphism, and could this relate to how they were able to expand their range so drastically?

In this study, individual genomes from populations of *V. berus* across Eurasia were sequenced to investigate how *V. berus berus* was able to colonize such a large range in comparison to other European vipers and further, why were they so successful compared to the other two *V. berus* subspecies. The Norwegian island population of *V. berus berus* will specifically be targeted for sampling to find genes associated with the rare phenotype and discuss the potential of this occurring in other populations of *V. b. berus*. This section will pose as an introduction to the background of *V. berus*, their phylogeography, the island of Gossa, similar cases of pattern polymorphism in European vipers, current research on pigmentation genes, aims and objectives of this study, the significance of this research and the possible limitations.

1.1 *Vipera berus*

The European adder, *Vipera berus* (family Viperidae), is one of the most widely distributed terrestrial snakes in the world. Its longitudinal distribution ranges between Sakhalin Island in the Far East of Russia to the Western coasts of France (Herreo & Fitze 2018; Lourdais et al 2013). In terms of its latitudinal distribution, *V. berus* can be found between northern Greece and above the Arctic Circle (Andersson 2003; Ursenbacher et al. 2006). Throughout this range, *V. berus* individuals commonly live in high altitudes (1,000-2,600 m above sea level; Ursenbacher et al. 2006) owing to a high cold tolerance that enables them to survive exposure to freezing temperatures for two to three hours (Andersson 2001). Brumation length, or the inactive period that ectotherm squamates experience during the colder months, differs across their range with activity periods as little as 17-18 weeks in northern populations (Andersson 2003).

Typically, males emerge in Spring, three weeks to one month before females to prepare for mating by basking in the sun, initiating spermatozoa production. After the first molt, males will actively search for females, only feeding after the year's mating season has ended (Nilson 1980). Conversely, females feed primarily before mating, reducing food intake to a minimum and in some cases entirely following conception. Females spend the remainder of the activity period intensely basking to expedite embryonic development (Lourdais et al 2013). This period of gestation depletes 30% of a breeding female's body mass leading to emaciation and in some cases death (Bauwens & Claus 2019), which is likely why they have adapted to a biennial breeding cycle while prioritizing hunting for non-breeding years (Nilson 1980).

Depending on availability, *V. berus* will feed on voles, lizards, amphibians, fish and birds, utilizing their characteristic dorsal zig-zag pattern for an "ambush hunter" strategy (i.e. waiting for prey in high activity zones; Bauwens & Claus 2019). Although this strategy is most typical of European adders, they also have been observed actively foraging (i.e. searching for potential prey), raising the question of whether this disruptive coloration is for hunting or to reduce predation (aposematism hypothesis; Santos et al. 2014) as is the case for other species of *Vipera* such as *Vipera latastei gaditana* (Bauwens & Claus 2019; Niskanen and Mappes 2005).

1.2 Phylogeography of *Vipera berus*

As periods of stadials and interstadials occurred throughout the Pleistocene, species like *V. berus* continuously shifted their range to seek out optimal habitat and, as a result, repeatedly recolonized Eurasia. In a study of mitochondrial DNA (mtDNA) based on a 1,043-bp region of the cytochrome *b* (COI) gene and 918 bp of the non-coding control region, three highly supported clades of *V. berus* were identified: an Italian clade including surrounding areas in Austria, Slovenia and Southern Switzerland; a Balkan clade containing *V. berus bosniensis*; and a Northern clade which stretched from the Far-East, westward to Great Britain and upward into Fennoscandia. The large range and sub-structure of the Northern clade suggested that two separate colonization events have taken place (Ursenbacher et al. 2006).

While there is not currently enough evidence to deduce whether the three clades (Northern, Balkan and Italian) diverged during a stadial or interstadial period of the first colonization, it is estimated to have occurred in the lower pleistocene. If they diverged in a stadial period, they were likely divided while seeking out colder temperatures, retreating to the Alps, Carpathian mountains or Balkan mountains. If this divergence occurred during an interstadial period, it may have been that they diverged somewhere in the lowland regions of the Carpathian, Balkan or northern Italian Peninsula. After the adders expanded from their refugia from the initial isolation event, the Northern clade dominated the largest range of the three while the Balkan and Italian clades were largely confined.

The second colonization event is estimated to have taken place as the climate shifted into an interstadial period. At this point, the Northern clade likely retreated to a number of refugia across the range from east of the Carpathians, to France or Hungary. Balkan and Italian clades traveled minimally, seeking favorable temperatures between high and low elevations determined by the glacial state. This reduction of distribution and lack of expansion into the Iberian Peninsula could be attributed to the presence of *Vipera seoanei* and *Vipera aspis* in the region (Ursenbacher et al. 2006a; 2006b) or due to a realized niche, considering that *V. berus* is more efficient at thermoregulation in cold climates (Lourdais et al 2013).

Suggested subclades within the Northern clade include the Carpathian, Eastern, Western and Central European. Basal positioning of the Carpathian subclade indicated that a Carpathian glacial refugium may have been the origin point of the Northern clade, which concurs with previous studies on the phylogeography of field voles (Ursenbacher et al. 2006; Jaarola & Searle 2002), frogs (Babik et al 2004) and newts (Wielstra et al 2017).

1.3 *Vipera berus* taxonomy

There are currently only three recognized subspecies of *V. berus*: *V. b. bosniensis*, *V. b. sachalinensis* and the nominate subspecies *V. b. berus*. The range of *V. b. berus* spans the entire distribution of *V. berus* with a low level of geographically correlated morphological variation (Cui et al 2016). The subspecies *V. b. sachalinensis*, named after Sakhalin Island (Russia), has a much smaller range and is less distinguishable from other subspecies in morphology and genetic differentiation, with some diagnostic characteristics being unobservable in some specimens (Cui et al. 2016). This is likely due to the scant sampling effort focused on it in combination with a lack of information about the subspecies. The proposed subspecies *V. b. nikolskii* was first described in 1986 as *Vipera nikolskii* but was later assumed to be a subspecies of *V. berus*. Currently it is not widely recognized as a subspecies of *V. berus*, although it has been described in detail with their habitat overlapping with the nominate subspecies in Ukraine (Zinenko 2006).

As a consequence of the migration events during the Pleistocene, the subspecies of *V. b. bosniensis* is confined to the Balkan region where it can be found from the lowlands of southwestern Hungary, Croatia and Northern Serbia to the Balkan and Carpathian mountains (Westerström et al. 2010; Ursenbacher et al. 2006a). In previous phylogenetic analysis using mtDNA (Ursenbacher et al. 2006b; Cui et al. 2016), the Balkan clade conforms well with *V. b. bosniensis* with high statistical support, clearly differentiated from the other subspecies, meaning the Balkan clade didn't manage to expand their range northward after the end of the last glacial period. A distinguishable difference between *V. b.*

berus and *V. b. bosniensis* is venom toxicity (Malina et al. 2011). The venom of *V. b. bosniensis* is unequivocally more potent with respect to neurotoxic activity than the venom of the nominate subspecies, resulting in very few records of neurotoxic symptoms in cases of *V. b. berus* envenomings (Malina et al. 2008). In regards to other taxa within the genus of *Vipera*, both *V. aspis* and *V. ammodytes* also produce neurotoxic venom (Malina et al. 2008; Zanetti et al. 2018). In addition, *V. aspis*, *V. ammodytes*, and *V. b. bosniensis* all inhabit southern Europe with *V. b. bosniensis* and *V. ammodytes* largely overlapping in the Balkan peninsula (Ursenbacher et al. 2008; Ursenbacher et al. 2006b).

1.4 Interesting morphology on Gossa

As mentioned, *V. berus* typically displays a striking dorsal zig-zag pattern ranging from gray and black to contrasting shades of brown depending on sex, excluding occasional cases of melanism in Germany, Austria, Netherlands, Sweden and Central Asia (Andren & Nilson 1981; Cui et al. 2016). Similar to other molting reptiles, vibrancy of coloration is also often subject to change depending on sexual maturity and time of the last molt (Smith 1951). Deviations from this morphology are extremely rare for the European adder, estimated to be as little as 2% of individuals (Hills et al. 2020). Rare sightings of adders with straight-lined dorsal patterns have been recorded in England and South Wales, but many displayed the disruptive zig-zag pattern elsewhere along the vertebrae with dorsolateral spots/bars. In a record of observations between 2005 and 2020, only two adders were reported to have a straight dorsal line without some form of disruptive patterning. This morphology had previously never been seen in high frequency, much less studied, until now.

On an island off the coast of Norway, a population of *V. berus* was recorded displaying this rare phenotype in approximately 10% of individuals during a preliminary analysis in 2017 (Martin et al. 2018, unpublished personal report; Ophus 2017). Using whole-genome sequencing, eleven individuals from Gossa and two from mainland Norway were extracted and sequenced, developing low-depth (3-11x) nuclear genomic data. Individual ancestry was inferred using NGSadmix, revealing that the island population was genetically distinct from mainland individuals. To further understand the basis of the stripe morph, two pregnant striped adders were collected and observed for the final weeks of the gestational period which resulted in 14 neonates, three being striped females and two being striped males. Although this only included two generations, it suggests that the stripe follows Mendelian inheritance.

Populations of *V. berus* such as these which were reported to have deviations from the zig-zag stripe were also devoid of melanistic individuals, possibly due to the ecological pressures from predators (Wolf & Werner 1994). Melanistic individuals, although more efficient at thermoregulation and consequently capable of growing larger from longer hunting hours, are more susceptible to predation (Andren & Nilson 1981). Snakes with longitudinal stripes are presumed to be more efficient using a rapid escape strategy against biotic dangers. Therefore, it may be that the adders on the island of Gossa face higher predation, posing melanism as a dangerous strategy and longitudinal stripes a more advantageous morph (Wolf & Werner 1994).

Gossa is an island roughly 50 km² in area that is covered in boreal heathland, moorland and marsh, with man-made strips of birch, sitka spruce and pine planted for wind

protection. Several locations have rocky outcrops with crevices that go deep into the earth. The lack of shade makes these South-facing hillsides perfect for wintering sites but vulnerable locations to bask. The landscape is speckled with numerous ponds and streams full of frogs (*Rana temporaria*) and fish, which the vipers regularly hunt using their tail as a lure (Kjeld Henrik Ophus, personal report). Other animals on the island include grazers such as deer, cattle and sheep as well as many coastal birds that use the island as a nesting and feeding ground when the adders are emerging in Spring. These coastal birds, as well as domestic cats, are the primary predators to the adders on Gossa.

1.5 *Vipera aspis* and the pro-opiomelanocortin (POMC) gene

Vipera aspis is a widespread snake indigenous to central and southern France, Switzerland and Italy. The range of *V. aspis* is second largest of the European *Vipera* genus and slightly overlaps with the southern range of *V. berus* (Zanetti et al. 2018). Populations of *V. aspis* deviate from a cryptic pattern across their distribution, exhibiting melanism in nearly 50% of individuals where the habitat is humid or high elevation (Ducrest et al. 2014). More rarely, few localized montane populations contain up to 50% concolor (uniform coloration) and striped morph vipers. In an analysis of the POMC and the melanocortin-1-receptor (MC1R) genes, a SNP within the 3'-untranslated region (3'UTR) of the POMC gene was discovered to be correlated with the dorsal color polymorphism in *V. aspis*. The correlation between the POMC gene and color morphology had previously been studied in humans and the famous agouti mice, with the loss of POMC-derived peptide α -MSH having a critical role in MC1R, causing red hair in humans and yellow coats in mice (Krude et al. 1998). Humans with mutations in the POMC 3'UTR have also been linked to having higher obesity rates, drug dependence and alcoholism (Sutton et al. 2005; Zhang et al. 2009). In fact, POMC impacts all melanocortin receptors in the melanocortin system by contributing to the production of adrenocorticotrophin (ACTH), melanocyte-stimulating hormones (MSH) α , β , and γ in addition to the opioid-receptor ligand β -endorphin. Mutations in the 3'UTR could lead to alterations in the post-transcriptional regulation process, causing mRNA degradation or repression of translation initiation (Ducrest et al. 2014). As POMC is a precursor to several hormones, it is consequently responsible for several vital functions in the body and is highly conserved across multicellular animals (Cai & Hruby 2016).

There are five melanocortin receptors: MC1R, MC2R, MC3R, MC4R and MC5R. The melanocortin receptors may be found in the adrenal gland (MC2R), brain (MC3R, MC4R) and throughout the body (MC5R). These receptors use POMC derived hormones to influence vital mechanisms for feeding behavior, energy homeostasis, stress response, UV radiation response, sexual function, pain response, fear flight, thermoregulation, etc. (Dores et al. 2016; Cai & Hruby 2016). Therefore, mutations within POMC can lead to drastic changes in the function of an animal as it has several pleiotropic effects. The MC1R is expressed in melanocytes and melanoma cells, with the increase in activity producing eumelanin and the decrease in activity producing pheomelanin (Rosenblum et al. 2004). Eumelanin and pheomelanin, responsible for brown/black and red/yellow pigments respectively, are both biosynthesized in melanocytes within the epidermis.

1.6 Pigmentation and pattern formation

As the study with montane populations of *Vipera aspis* and similar studies on other squamates (Rosenblum et al. 2004; Corso et al. 2012; Cox et al. 2013) did not find correlation between color polymorphism and the MC1R, it is imperative to note that pigmentation in snakes differ from that of mammals due to an additional four pigmentation cells called chromatophores. With the exception of iguanas and various chameleons, the stratification of chromatophores begins with melanophores at the base of the dermis producing black pigment patterns. Ascending from the melanophores are the iridophores, erythrophores and xanthophores which are located in the loose dermis, causing a light scattering effect and red/yellow pigments respectively. The pattern diversity of snakes is the result of the chromatophores and melanocytes in these layers of the dermis and epidermis interacting (Ullate-Agote et al. 2020; Alibardi 2013; Saenko et al. 2013).

Another factor of pattern morphology is determined by the ability of pigment to bind to vesicles in the melanocytes and chromatophores. To achieve this, an unpigmented step must take place to lay the foundation for melanin polymerization, condensation and storage. In melanocytes, processing of pre-melanosomal protein (PMEL) initiates the formation of intraluminal fibrils which creates a scaffold for eumelanin. In experiments inactivating PMEL in mice, a mild discoloration was observed, giving the mice lighter coat and skin color, but it led to a significant reduction of eumelanin in hair (Bissig et al. 2016; Hellström et al. 2011). Similarly to melanosomes, xanthosomes also display fibrous material for the pigment composition. Iridophores, however, create guanine crystal pigments in their platelets and organize them in dense arrangements. Such arrangements secure iridophores as the major contributor to striped patterning in zebrafish with experiments of induced iridophore deficiencies causing pattern defects such as alterations in stripe positioning and loss of stripes (Ullate-Agote et al. 2020; Gur et al. 2020).

1.7 Experimental design

As stated, melanin-based polymorphism has been well studied in mammals and birds with the research primarily based on melanocytes, which has consequently led to research on the color polymorphism of distantly related animals being confined to POMC, ASIP, SILV/PMEL, and the MC1R. Zebrafish have had a significant role in the understanding of chromatophores and pattern formation as the model species for vertebrates but have resulted in limited research on the complex color and pattern morphology of reptiles (KURIYAMA et al. 2020). Therefore, the mechanisms in which color polymorphism is developed in reptiles had often been theorized (Chang et al. 2009; MURRAY 2011; Murray & Myerscough 1991; Allen et al. 2013) but only recently studied in vivo. Using next-generation sequencing or whole-genome sequencing in an analysis of color polymorphism is a relatively unbiased approach that broadens the search for relevant genes rather than restricting them to known markers used in previous studies of mammals and birds (KURIYAMA et al. 2020). This approach is also beneficial for reptiles due to the lack of representation in the sequencing databases. With projects like the "Reptile Transcriptomes Database 2.0", this representation has increased dramatically, but with there being nearly 10,000 species of squamates, the database is continuously expanding as more sequences become available

(Tzika et al. 2015). As of April 2022, there was no sequence available on the database within the subfamily of Viperinae.

Given the lack of research on reptile color polymorphism in regards to pigmentation genes outside of melanocytes, as well as the lack of reptile representation on sequencing databases, this project aimed to determine the genomic basis of the rare phenotype on the island of Gossa using whole-genome sequencing to investigate why Gossa has this phenotype and which genes are associated with it. Additionally, subspecies of *V. berus* and populations across Eurasia were selected for sampling to represent the entire range of the Northern clade from the three inferred Pleistocene refugia and to investigate why *V. berus berus* successfully colonized its broad range while other *V. berus* subspecies stayed localized to their Pleistocene glacial refugia. Following various bioinformatic approaches, the phylogeny of the European adder was inferred, genes associated with the stripe morph were identified and the potential for other European populations to develop this morph could be discussed. The findings in this study will contribute to the understanding of pattern development and pattern polymorphism in wild populations of snakes by taking the relatively unbiased approach of using genome scanning. With that, possible limitations of this approach lay heavily on the genome annotation available for *V. berus*. Late into the creation of this project, the realization that the GenBank reference genome was not annotated forced the use of an *ab-initio* genome annotation created using MAKER v3.01.03, SNAP and Augustus (Cantarel et al. 2008; Korf 2004; STANKE et al. 2006) resulting in 7,948 gene models, which falls short from the estimated 21,194 predicted protein-coding genes found in *Deinagkistrodon acutus*, a fellow member of the family Viperidae (Yin et al. 2016). Additional limitations restricting the power of this analysis were sample size and the sample quality of archival tissue samples.

In this section, the background of this experiment has been outlined with clear study objectives. The next section will cover the methodology in detail to ensure reproducible results, and the section following will reveal the results of the experiment with figures. Lastly, a section discussing the application of these findings in context of the field's current research will wrap up the paper.

Methods

2.1. Sample acquisition

The outgroup taxa used in the primary analyses included five genus of *Vipera* as well as three subspecies of *V. berus*. As the ingroup, there were seventeen populations of *Vipera berus* spanning twelve countries across Eurasia to ensure representation of the three inferred Pleistocene refugia (Ursenbacher et al. 2006). Blood, buccal swabs, and tissue samples (both fresh and archival) were obtained through collaborators, combining freshly harvested samples and museum collections. The Norwegian samples were all freshly harvested.

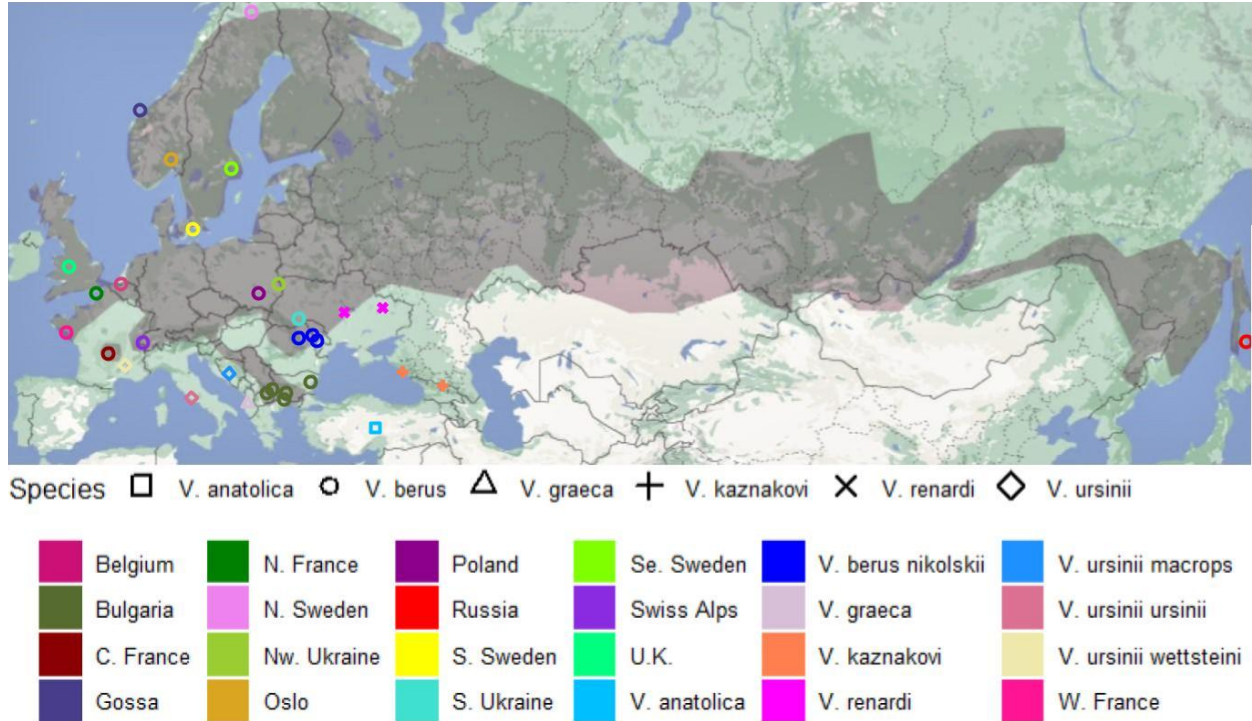


Figure 2.1.1: Map of sampling locations. The shaded area displays the range of *Vipera berus* while the points represent sampling locations.

Sampling on Gossa

78 snakes were captured and sampled on the island of Gossa, 58 being non-striped and 20 being striped. Gossa was heavily sampled to better represent the populations on the island as well as to find genes under selection regarding the rare morphology. The expedition took place at the beginning of April during the first weeks of warm weather. As these were the first snakes to come out of hibernation, most individuals collected were male.

Ideal conditions on the island were spotted, which included south-facing rocky outcrops, where *V. berus berus* naturally hibernates. After capture, each snake was photographed and identified using their head scalation pattern and marked with water-soluble, non-toxic marker. Following this, the individual's SVL (snout-vent length), tail length, weight and sex were recorded. DNA samples were collected using two methods to ensure successful extraction of each individual. Buccal swabs were collected according to Boca Scientific instructions (SF 1), which entailed leaving the buccal swab in the mouth of the snake for approximately 30 seconds, then placing it into the collection tube with a silica capsule that quickly dries the sample to stabilize the DNA. Ventral scale clips were taken from each individual using tweezers and scissors. Each clipping was then placed into 95% ethanol to stabilize it. These methods of DNA collection were chosen because they are minimally invasive for live snakes.

2.2 DNA extraction, library preparation and sequencing

Genomic DNA was extracted from buccal swabs, blood and scale clippings with the Qiagen DNeasy Blood and Tissue extraction kit following the DNeasy® Blood & Tissue Handbook (SI 1) using the spin column method. After, the DNA was sheared using the Covaris or Bioruptor sonication instruments. The shotgun genomic libraries were constructed using the single-tube, optimized, *BEST 2.0 library build for 96 samples in well plate format v. 1 (Blunt-End Single-Tube; 2019)* protocol based on the method employed in Carøe et al. (2018; SI 2). High-throughput sequencing was performed at the Novogene-Europe sequencing facility.

2.3 Bioinformatic analysis

The Paleomix pipeline v1.3.6 was chosen due to its efficiency in handling datasets that originate from a combination of modern and degraded museum collection samples (Schubert et al. 2014). The Paleomix pipeline is a collection of bioinformatic tools designed to, among other things, automate the alignment of de-multiplexed sequence reads against a reference sequence. This process began by using AdapterRemoval v2.3.2 to trim low quality and ambiguous bases from the 3' end. The sequence reads were then mapped against a *Vipera berus* reference genome (Genome assembly Vber.be_1.0; NCBI accession code GCA_000800605.1; Clark et al. 2016) using the Burrows Wheeler Aligner (BWA) v 0.7.16a with the backtrack algorithm. There was allowed a minimum map quality of 30, and the unmapped reads as well as the PCR duplicates were filtered. The BAM files generated by the Paleomix pipeline were filtered by sequencing depths after assessing their summary files. Individuals with mean sequencing depths below 0.1 were removed from the dataset entirely.

The programs RepeatMasker version 10.0.11 (Smit et al. 2015) and RepeatModeler version 2.0.3 were paired to screen the DNA for interspersed repeats and low complexity DNA sequences, flag and mask them. The output from RepeatMasker then had the repeats removed to be used as an input in the software package ANGSD (Analysis of Next Generation Sequencing Data) version 0.934 (Korneliussen et al. 2014). This program suite was used to parse the BAM files using several error models to calculate genotype likelihood (GL). This program creates a beagle file of GLs based on the aligned reads, association mapping and sequencing quality scores using the GATK model. This was vital in this dataset primarily due to low coverage data acquired from museum samples being error prone (Hahn et al. 2018). For downstream analyses, the GLs will be taken into account, incorporating statistical uncertainty. The (Genome Analysis Toolkit) GATK model uses the MapReduce philosophy which optimizes the efficiency of processing large datasets (Dean & Ghemawat 2008; McKenna et al. 2010). To make the beagle file to be used in later Plink analysis, the major and minor alleles were inferred from the likelihoods using the `-doMajorMinor 1` flag in ANGSD. The `-doPlink` and `-SNP_pval` were also used to produce files for the Plink analysis by producing a TPED file by doing a pseudo haploid call for sites and setting a cutoff of P-values more than $1e-6$, respectively. Flags `-doGeno 1`, `-doPost 1`, were used to print out the major and minor from genotype calling and calculate posterior probability using the frequency as a prior. To account for mapping and base quality, the flags `-minMapQ` and `-minQ` were used. The minimum sequencing depth to be called was set using the `-setMinDepthInd`, genotypes called were filtered using the `-postCutoff 0.95`. The run was also set to only use the sites that were covered in half the number of samples using the `-minInd` flag. PCR duplicates and

sites with multiple best hits were removed using the flags `-remove_bads` and `-uniqueOnly` flags.

2.4 Population structure and phylogenetics

The toolset PLINK 1.90 (Purcell et al. 2007) was used to calculate pairwise linkage disequilibrium (LD) between SNPS. The average LD distances were plotted to visualize the decay of r^2 in LD and the threshold for pruning was determined by taking half of the initial r^2 value (Hahn et al. 2018). To estimate the covariance matrix and allele frequencies, the remaining variants were analyzed by PCAngsd v.1.03. PCAngsd uses the genotype likelihoods from heterogeneous populations to infer population structure using principal component analysis (PCA) on low-depth NGS data. The covariance matrix was then loaded into R v.4.0.5 to calculate the eigenvectors and eigenvalues using the `eigen` function included in the base R v.4.0.5. Two PCA plots were created with GGplot2 v.3.3.5 (Wickham 2016) using the eigenvectors with ellipses based on 95% confidence intervals. One plot included all species and subspecies (*V. berus berus* and populations of *V. berus bosniensis*, *V. renardi*, *V. anatolica*, *V. graeca*, *V. kaznakovi*, *V. ursinii macrops*, *V. ursinii ursinii*, and *V. ursinii wettsteini*), the other included only the populations of *V. berus berus*.

To begin estimating the pairwise genetic distances, the program `ngsDist` was used, which uses the GLs instead of the genotype calling. This approach factors in the statistical uncertainty of low-depth sequences which increases the accuracy of the estimates for tree building and bootstrap support on phylogenetic trees (Vieira et al. 2016). Using the software `Fastme 2.1.4` (Lefort et al. 2015), a tree was constructed from the pairwise genetic distances using the `BalME` criterion by selecting the flags `-m b` and `-n b`. On a second run of `ngsDist`, the branch support values were created using `-n_boot_rep` to produce 100 bootstrap replicate matrices and a block size of 20 with `-boot_block_size`. Using `RAXML v8.2`, the branch support values were placed on the final tree using the `GTRCAT` approximation and drawing bipartition information on the tree (Stamatakis 2016).

2.5 F_{ST} Analysis

The F_{ST} analysis was separated into three sections: striped individuals vs. non-striped individuals on the island of Gossa, pairwise comparisons of *V. berus berus* across Eurasia, and *V. berus berus* vs. *V. berus bosniensis*.

1D SFS (site frequency spectrum) values were calculated for both the striped ($n=19$) and unstriped ($n=58$) sub-populations using `RealSFS` in `ANGSD` version 0.934 (Korneliussen et al. 2014). To estimate the SFS, the site allele frequency likelihood was calculated using the GLs with the flag `-doSaf 1`. The *V. b. berus* reference genome, containing 28,883 scaffolds, was used in place of the ancestral state which was assigned using the `-anc` flag. The positioning of each variant was inferred using the scaffolds. The GATK model was implemented using the `-GL 2` option and the quality threshold was set using `-minMapQ 25` to only use reads with a mapping quality above 25. Further filtering was accomplished by setting a minimum quality score of 20 using `-minQ 20` and removing reads above 225 with `-remove_bads 1` (duplicate reads) and discarding sites with multiple best hits using `-uniqueOnly 1`. Using the striped and unstriped 1D SFS, the 2D SFS was calculated. The `-fold` option was included to estimate the folded SFS since there was no ancestral state

used. The F_{ST} index was calculated from the SAFs from both striped and unstriped sub-populations as well as the folded 2D SFS using the Hudson estimation method in realSFS with the `-whichFST 1` option. The realSFS F_{ST} stats were then used to find the weighted and unweighted F_{ST} values. Using the realSFS F_{ST} `stat2` function, "sliding windows" 10-kbp long "slid" every 10-kbp using `-win 10000` and `-step 10000`. This method was chosen as it considers observations in the region of the genome instead of independent SNPs (Beissinger et al. 2015) to improve the probability of finding the genomic basis of the stripe phenotype.

The results from the F_{ST} analysis were cleaned by removing negative F_{ST} values, windows with more than 100 SNPs and windows with very few SNPs. To identify outliers and potential genes of interest, the z-scores were calculated and ranked using the quantile function in base R. $Z\text{-score} > 15$ were highlighted on the plot created with GGplot2 v.3.3.5 (Wickham 2016). These outlier regions were extracted from the *V. b. berus* reference genome and searched using the Basic Local Alignment Search Tool (BLAST) on NCBI to check if any genes were associated with the sequence. By searching for these regions on the annotated genome, the genes in the region were identified.

The Eurasian populations of *V. b. berus* were inferred from the principal component analysis and reduced to only include populations with $15 > n > 5$. A subset of 15 individuals was randomly selected from the Gossa population to include 6 striped and 9 unstriped individuals. The comparisons of *V. b. berus* populations began similarly with 1D SFS values calculated for each population ($n=11$) followed by 2D pairwise SFS estimates between each population pair (55 comparisons). Using the realSFS F_{ST} index and realSFS F_{ST} `stats2` function, the final weighted and unweighted F_{ST} values were calculated and converted to a heatmap with the `pheatmap` v.1.0.12 R package. Additionally, a distance matrix of geographic distances was created by first calculating a general geographic coordinate for each population and then using the `dist` function in R v.4.0.5. The isolation by distance and correlation by distance was calculated with the `mantel.rtest` function using the spearman method, which was plotted using GGplot2 v.3.3.5 (Wickham 2016). To further visualize the isolation by distance, a regression of the genetic distances ($F_{ST}/(1 - F_{ST})$) and geographic distances (km) was also produced using the spearman method.

The final F_{ST} calculation included 60 *V. b. berus* and seven *V. b. bosniensis* individuals. The *V. b. berus* subset contained five randomly selected individuals from each of the populations in Belgium, Southern Sweden, Northern Sweden, Eastern Sweden, Switzerland, Central France, Northern France, Western France and Eastern Russia. Four snakes from the United Kingdom and one from Poland were also included. The 1D SFS and 2D SFS estimates were calculated as in the biogeographic comparison. Further, the realSFS F_{ST} index and F_{ST} `stats` function were used to obtain weighted pairwise F_{ST} values. The parameters set for the striped vs. unstriped analyses were repeated for all functions here as well as for the "sliding window" method. The F_{ST} results were cleaned by removing negative F_{ST} values, windows with more than 100 SNPs and windows with few SNPs. The F_{ST} scores were z-transformed using the same methods. After analyzing the plot from GGplot2 v.3.3.5 (Wickham 2016), windows with $z\text{-score} > 5.5$ were selected as regions of interest and searched on BLAST as well as on the annotated *V. b. berus* genome to identify genes in the region.

2.6 Genome-wide association study (GWAS)

The software ANGSD was used to produce a .beagle file of GLs (-doGlf 2) using the GATK model (-GL 2). Major and minor alleles were inferred directly from the GLs using the maximum likelihood approach (-doMajorMinor 1) with a minimum mapping quality limit of 25 (-minMapQ 25) and minimum base quality of 20 (-minQ 20). The alleles were written in the tfam/tped format (-doPlink). Posterior probabilities above 0.95 (-postCutoff 0.95) were calculated using frequency as the prior (-doPost 1) to be used in the creation of .maf files (-doMaf 3), excluding minor alleles with low frequency (-minMaf 0.05) since imputation using BEAGLE drops in accuracy for rare alleles with a minor allele frequency <0.05. The minimum sequencing depth was determined (-doCounts 1) using the frequency of bases (-doCounts 1). SNPs with p-values over $1e^{-6}$ were removed (-SNP_pval $1e^{-6}$) as well as PCR duplicates and sites with multiple best hits (-remove_bads 1 -uniqueOnly 1). Repeats were identified and flagged with programs RepeatMasker version 10.0.11 (Smit et al. 2015) and RepeatModeler version 2.0.3. The repeats were removed from the .bed file to create a clean .beagle file. As the filtering of SNPs in linkage disequilibrium is particularly important for GWASs, these were detected using PLINK 1.90 (Christoforou et al. 2012; Purcell et al. 2007). Bins of the pairwise comparisons were calculated using the LD distances and plotted in R to visualize the LD decay. The pruning threshold was determined by the initial r^2 value/2 (Hahn et al. 2018). Regions to be pruned were identified with windows of 25 kbp and the pairwise cut-off of r^2 value/2, jumping every 5 kbp. To produce a PCA from PCAngsd, the eigenvalues and eigenvectors were calculated from the covariance matrix. The plot was constructed using GGplot2 v.3.3.5 (Wickham 2016) and the PC eigenvalues were recorded for the GWAS analysis.

Genotypes were imputed using BEAGLE 3.3.2 (Stahl et al. 2021) with an initial 3,530,627 SNPs and 154 haplotypes. Following imputation, LD was pruned again with the previously used parameters generated from PLINK since LD affects the GWAS heritability (GWASh) estimates. Using ANGSD, the probability file from the imputation (-yBin), and the covariates file (-cov), the parameter of heritability estimates and genome wide association was calculated. The model chosen was the latent genotype model (Expectation maximization algorithm; -doAsso 4), which uses the genotype probabilities to estimate heritability. The allele frequency was similarly estimated using the genotype probabilities (-doMaf 4). The minimum number of highly credible genotypes (-minHigh) and minimum number of minor alleles (MAF; -minCount) were limited to 10. Instead of likelihood ratio statistics, the p-value (-Pvalue) was calculated to be used for plotting results. Finally, the -fai flag pointed to the reference genome index to be used in the association.

The GWAS association test was run on Computerome 2.0 with the aid of Dr. Jaelle Brealey. This association began by loading htlib/1.13 and angsd/0.935 for another ANGSD run. As before, genotype probabilities and phenotypes were supplied for the run in addition to the *V. berus* reference genome. The flags from the previous ANGSD run were also retained for this analysis with the exception of the -minHigh and -minCount, which were both increased to 30 due to the low number of Gossa individuals (n=77) in comparison to larger GWAS. Also, the first six principal components from the PCA were included as covariates to reduce genomic inflation and correct for the population structure caused by two groups of highly related individuals, similarly to the method used by (Batai et al. 2021). SNPs that ANGSD filtered out were removed from the results. Finally, the association was

explored using R packages *reshape2*, *GGplot2* v.3.3.5 (Wickham 2016) and *qqman* to visualize genomic inflation and haplotypes with high association. False discovery rate (FDR) estimates were calculated using the *RAINBOWR* R package (Hamazaki & Iwata 2020) with a significance level of $\alpha = 0.01$. Both the Bonferroni adjustment and Benjamini-Hochberg procedure methods were used to identify significant markers on the Manhattan plot. Using the genome annotation and *bedtools intersect*, each significant position was searched for genes that intersected to determine genes associated with pattern variation on Gossa.

Results

The resulting sequences from this mix of modern and degraded museum collection samples, produced on average 2.87x coverage when mapped against the European adder genome. Only individuals with >0.10x coverage (n=227) were included in the principal components analysis, while individuals with >0.50x coverage (n=201) were used in the phylogenetic and F_{ST} analyses. Pruning the variants in high linkage disequilibrium from the sampleset with >0.10x coverage resulted in 425,019 SNPs for the analysis of population structure. The sampleset with >0.50x coverage was pruned separately, yielding 489,962 SNPs for the other analyses mentioned above. The Gossa samples with >0.50x coverage (n=77) were independently pruned for LD before the GWAS, which included the use of 370,378 SNPs.

3.1 Population structure and phylogeny

The principal components analysis (Figure 3.1.1) differentiated *V. berus* from the other species of *Vipera* along PC1, which could explain 26.7% of the variance in the dataset, although *V. berus* populations from Russia, Bulgaria and Ukraine displayed more similarity to these outgroup species than *V. berus* populations in Central, Western and Northern Europe. This likely is due to the majority of individuals representing these countries being subspecies of *V. berus*. Variation along PC2 separates the geographic origins of various *V. berus* populations, with the French and Swiss clusters being highly differentiated from other countries and each other. Northern European countries were distributed opposite of the French and Swiss populations along PC2, with Norway and Sweden largely overlapping. Variation within the Norwegian and Swedish populations are displayed along PC1. This was also the case for Belgium and England, presumably remnants of the colonization event after the last Ice Age. PC3 and PC4 could explain only little variation between groups, with the exception of Scandinavia, northwest Ukraine, Russia, France and Switzerland. PC3 and PC4 also appear to divide the Northern clade as suggested by (Ursenbacher et al. 2006) into the Eastern, Western, Central European and a Far East subclade although the strength of these eigenvalues are low (< 0.1).

This distinction between populations in the Northern clade is also evident in the PCA of *V. berus* subspecies (Figure 3.1.2) with PC1 largely differentiating the Far East, Eastern, Carpathian and Western subclades. The distribution of the Carpathian subclade (Southern Ukraine, Poland and *V. b. nikolskii*) along PC2 divides populations while PC1 displays the variation within each Carpathian population. The Central and Western subclades are distributed along the secondary PC, with Scandinavia being highly differentiated from the Western subclade (Switzerland and France). Belgium and the U.K. fall between these groups

along the same axis, again, overlapping because of their recent separation after the last glaciation. PC3 stratifies the Western subclade with Northern France and Switzerland on opposite poles. This variable has also differentiated the Central-European subclade (Belgium and the U.K.) from the remaining populations. PC4 is less informative (eigenvalue <0.05) but has separated the Scandinavian countries as well as the Far East and Eastern subclades.

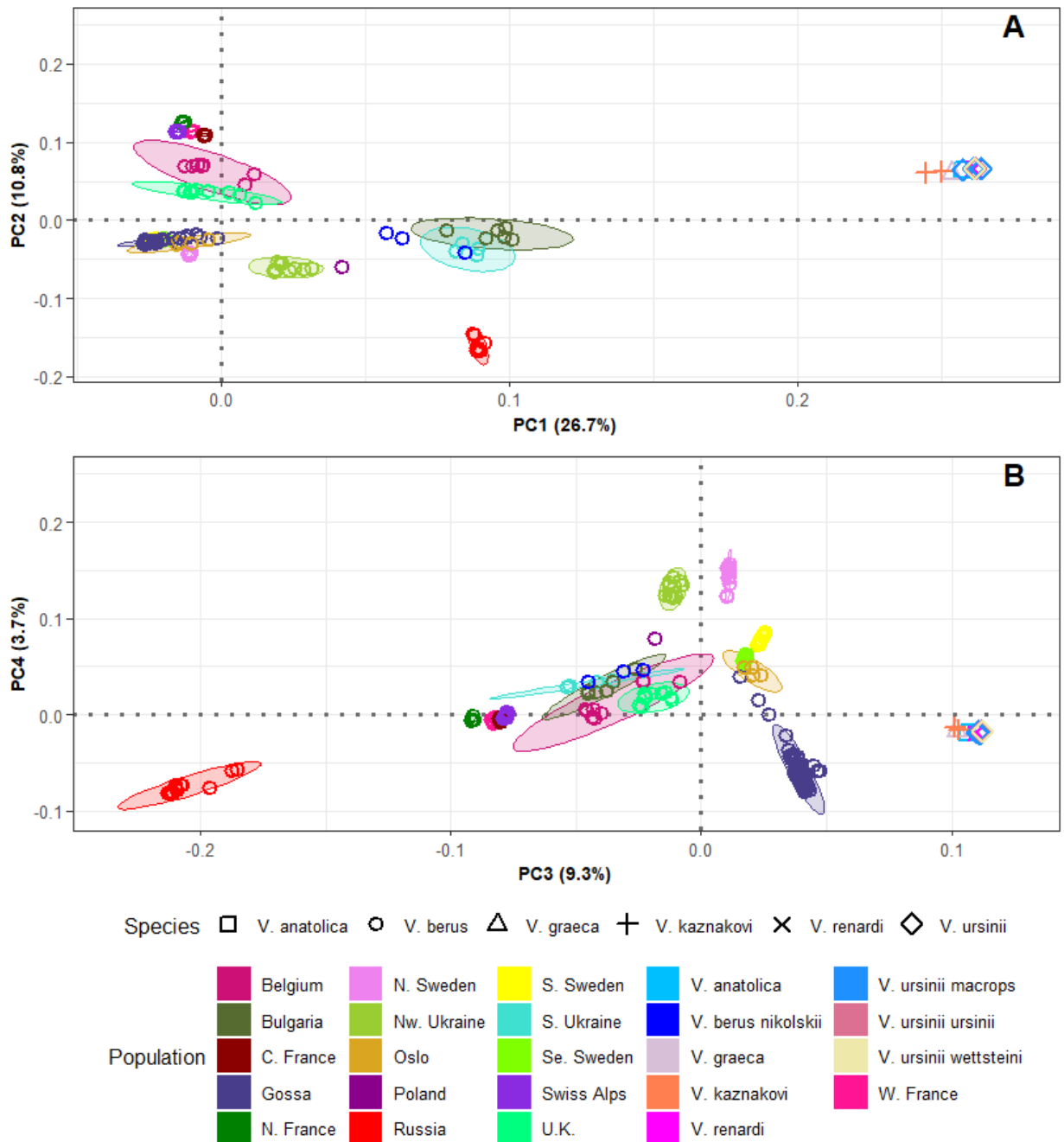


Figure 3.1.1: Principal component analysis exhibiting the population structure of all samples (n=227) with the first three principal components capturing > 46% of total variance. Eigenvectors and eigenvalues were generated from the genetic covariance matrix

obtained from the PCAngsd analysis. Ellipses were based on 95% confidence intervals of the populations separated by species and population. Ellipses were not estimated for countries with too few samples, such as countries representing outgroup species of *Vipera*. Also, subspecies of *V. berus* have been combined to represent the nominate species. Similarly, subspecies of *V. ursinii* were combined to represent all of *V. ursinii*. Colors correspond with the sampling locations in Figure 2.1.1. **A)** PC1 (26.7%) vs. PC2 (10.8%). **B)** PC3 (9.3%) vs. PC4 (3.7%).

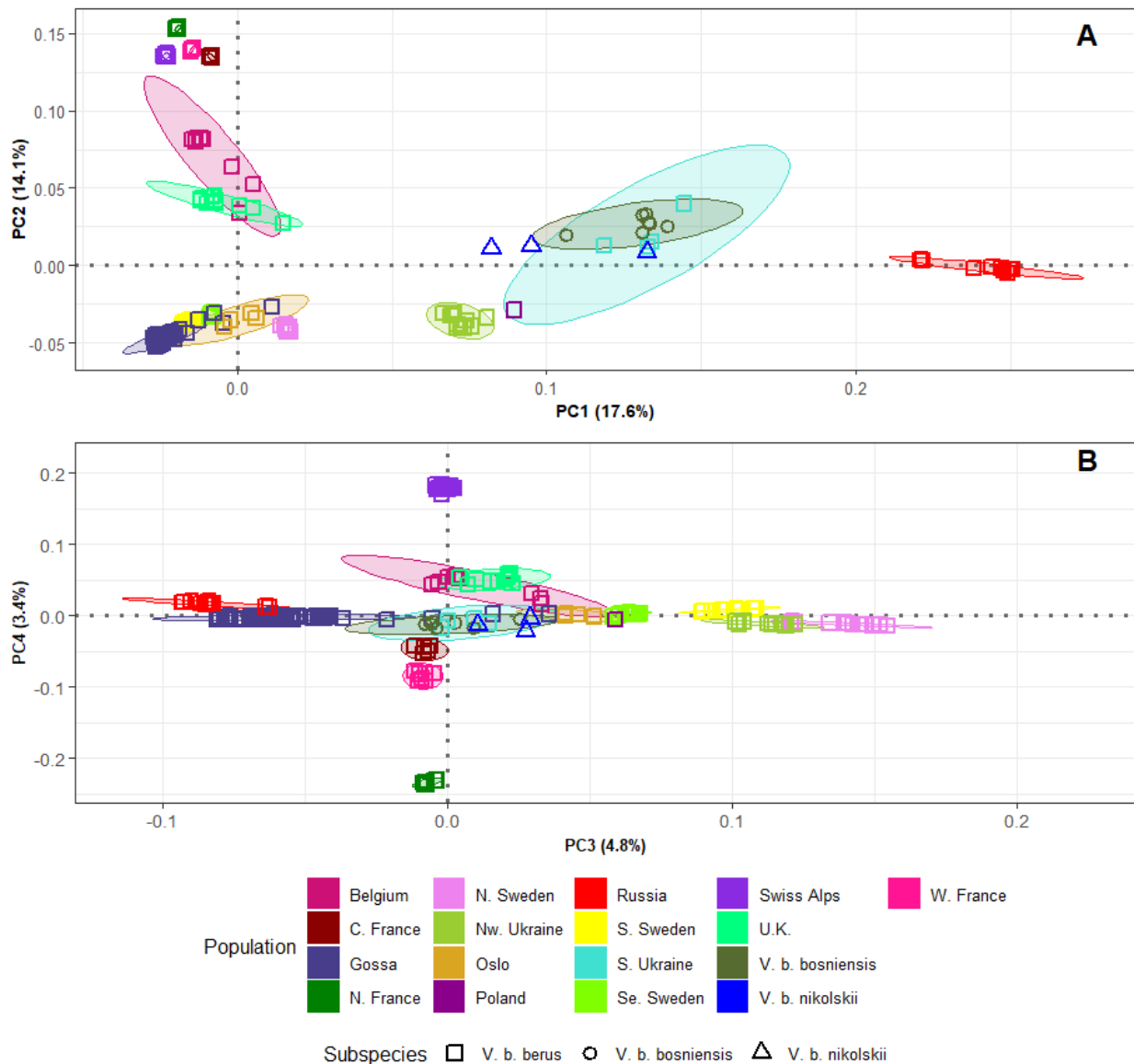


Figure 3.1.2: Principal component analysis exhibiting the population structure of *V. berus* and subspecies (n=201) with the first three principal components capturing >36% of total variance. Eigenvectors and eigenvalues were generated from the genetic covariance matrix obtained from the PCAngsd analysis. Ellipses were based on 95% confidence intervals of the populations separated by subspecies and sampled population. Ellipses were not estimated for populations with too few samples, such as Poland and *V. b. nikolskii*. **A)** PC1 (17.6%)

and PC2 (14.1%) of the principal component analysis. Populations of *V. b. berus* were identified using the calculated ellipses to be used in the F_{ST} analyses. The resulting populations were Belgium, Southern Sweden, Northern Sweden, South-east Sweden, Swiss Alps, Central France, Northern France, Western France, Russia, Gossa and Ukraine. The population in Oslo was excluded from this list as there were too few samples. **B)** PC3(4.8%) and PC4(3.4%) of the principal component analysis. Colors correspond with the sampling locations in Figure 2.1.1.

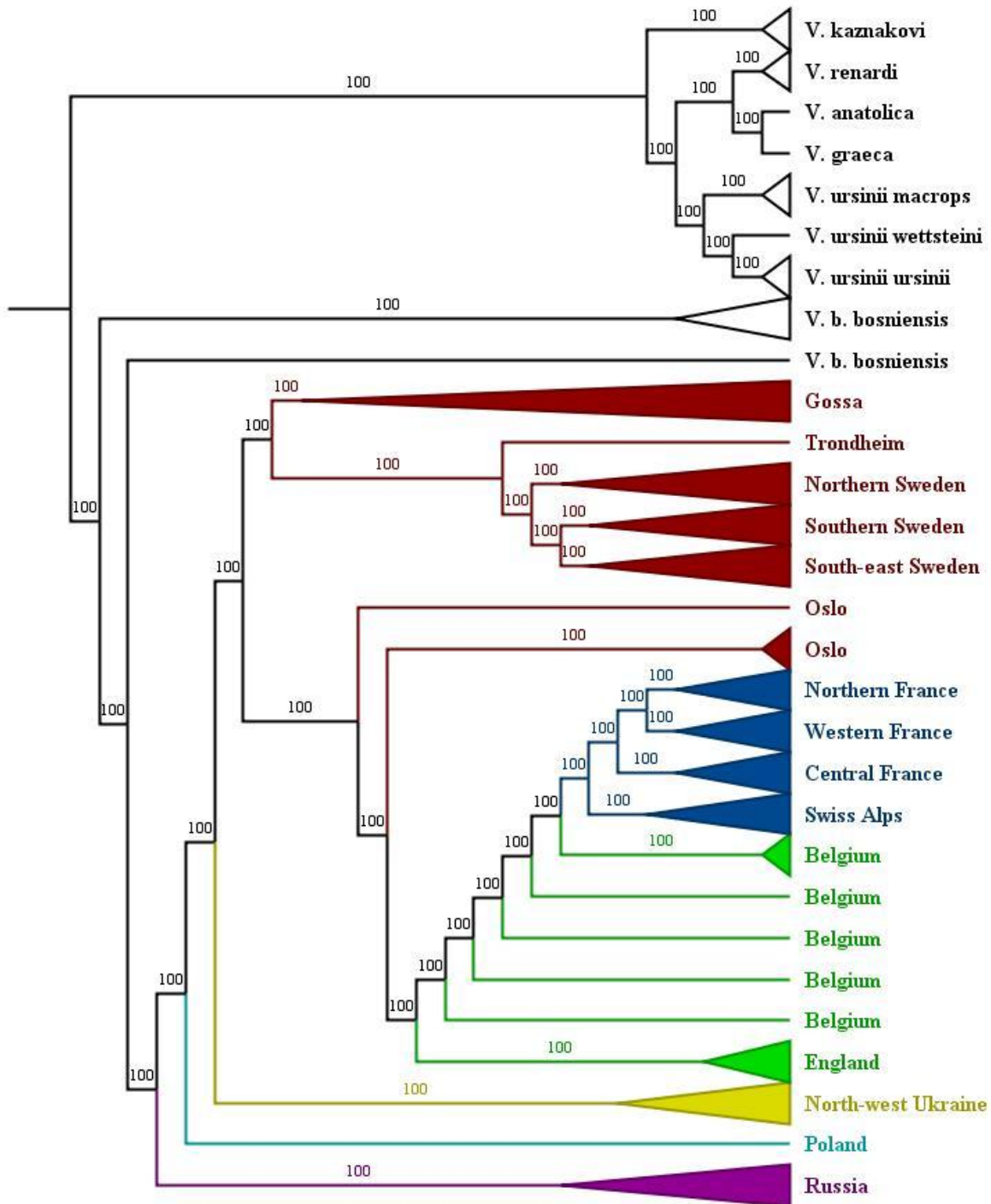


Figure 3.1.4: The maximum-likelihood phylogenetic tree of the Northern Pleistocene refugial clade (Ursenbacher et al. 2006a; 2006b). Outgroup species of *Vipera* are located at the top of the tree which has been selected as the root of the tree. *Vipera berus* subspecies: *V. b. bosniensis* has been used as the primary outgroup for the Northern clade, as *V. b. bosniensis* represents the Balkan clade. The branches have been colored to represent the five inferred subclades from (Ursenbacher et al. 2006) with the addition of a Scandinavian group. Far East = Purple, Carpathian = Teal, Eastern = Yellow, Central European = Green, Western = blue, Scandinavian = red, outgroup species of *Vipera* = Black.

The ML reconstruction of the outgroup species of *Vipera* (Figure 3.1.4) revealed well-supported phylogenetic relationships. In the earliest diverging clade, *V. kaznakovi* is positioned as the basal lineage, sister to the remaining outgroup species. These remaining species form a secondary clade that consists of a monophyletic clade of *V. ursinii* and a clade including *V. renardi*, sister to *V. anatolica* and *V. graeca*. Within the *V. berus* clade, *V. b. bosniensis* diverged as the basal lineage to the remaining *V. berus* populations, which was expected as *V. b. bosniensis* was included in the analysis to represent the Balkan clade of the three inferred Pleistocene refugia (Ursenbacher et al. 2006a). The Northern clade can then be further divided into a series of nested clades: Sakhalin, Carpathian, Eastern and a paraphyletic clade containing the Central European group with a monophyletic Western crown group nested in it. The first clade, containing the Sakhalin population from the Far East, has yielded a monophyletic basal lineage to the remaining European clades, distinct from the Eastern clade which supports the differentiation between *V. b. sachalinensis* and the nominate subspecies. The third clade was the Carpathian, which diverged as the basal lineage to the remaining clades, as demonstrated by (Ursenbacher et al. 2006). Though this clade is undersampled, containing only one Polish sample, the isolation from the Eastern and Sakhalin groups on the PCA is representative of its distinction (Figure 3.1.1; Figure 3.1.2). The population in northern Ukraine formed the fourth clade (Eastern) by splitting from the Western, Central European and Scandinavian group. This group was paraphyletic due to the clustering of three Oslo samples in the Western and Central European cluster. The final monophyletic group was the Western clade, which was nested within the Central European group.

3.2 F_{ST} analyses

Striped vs. unstriped

The pairwise F_{ST} score calculated from the site frequency spectra of both sub-populations resulted in a low weighted F_{ST} of 0.008, which was to be expected as the entire island is considered to be a breeding population (Kjeld Henrik Ophus, personal communication). The sliding window analysis resulted in 123,963 windows, 124 of which were outliers with z-scores in the 99.9% quantile, represented by the red line (Figure 3.2.1). The outliers were further filtered to include only SNPs with z-scores >15 to isolate regions with exceptionally strong divergence, which generated a list of 15 regions (Table 3.2.1). Regions were then retrieved by extracting the outlier using the bp position and then scanning them for genes on the annotation. Three genes were identified using this tactic:

SUOX, DGKA and PMEL. Remaining outlier regions that did not contain genes in the *Vipera berus* annotation were characterized by performing BLAST searches using the extracted region sequences as queries. This method revealed DDN, ATF2, and the SVMP complex as potential genes of interest (Table 3.2.1), while taking into consideration query coverage, percent identity and e-values. No genes were successfully identified in window 35555 using either BLAST searching or the genome annotation.

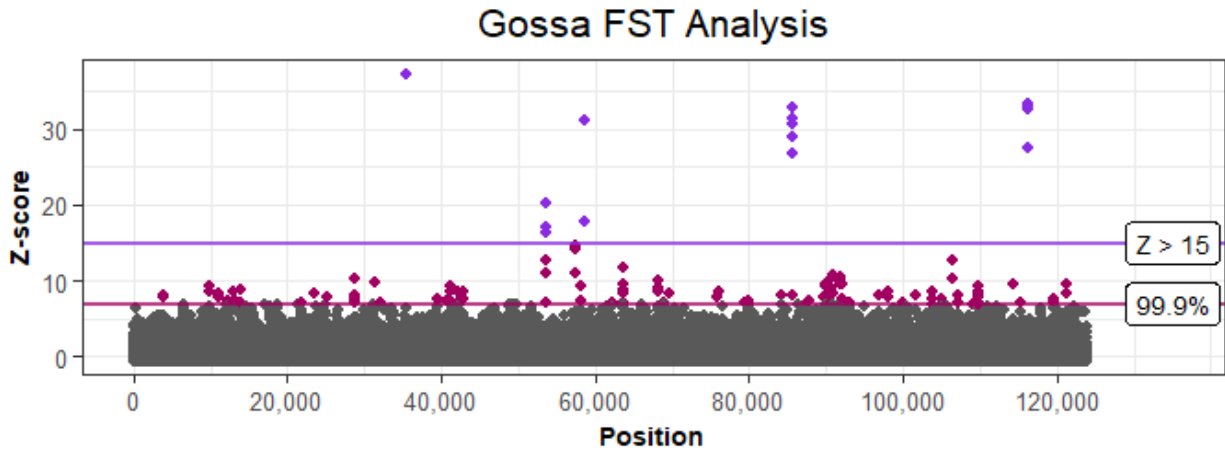


Figure 3.2.1: The plot of a pairwise F_{ST} sliding-window analysis for the Gossa population using 10 kbp windows and steps of 10 kbp. The red line at $y=6.98$ represents the threshold of the 99.9% quantile. Outliers with z-scores >15 are colored in purple as a subset of this quantile to identify regions of interest with noticeable peaks. The plotted dots are the midpoints of each window.

Outlier region				Gene identification					
Window number	Scaffold	Position	Z score	Gene(s)	Sites	Query coverage	E-Value	Identity	Accession
35555	KN623572.1	1,358-1,588	37.3	-	232	-	-	-	-
53621	KN631330.1	6,035-14,599	20.1	B-type SVMP	8,566	13%	2e-56	82.78%	MT032003.1
53622	KN631330.1	14,600-22,990	16.5	B-type SVMP	8,392	13%	2e-56	89.12%	MT032003.1
53623	KN631330.1	22,991-32,625	17.1	DDN	9,636	31%	0	94.62%	XM_039326753.1
58830	KN631677.1	8,036-17,895	17.7	ATF2	9,861	6%	1e-68	80.57%	KU866087.1
58831	KN631677.1	17,896-26,213	31.3	SUOX	8,319	-	-	-	-
85704	KN633923.1	8,814-17,645	32.8	DGKA	8,833	-	-	-	-
85705	KN633923.1	17,646-27,594	31.4	DGKA	9,950	-	-	-	-
85706	KN633923.1	27,595-37,592	26.8	DGKA, PMEL-like	9,999	-	-	-	-
85708	KN633923.1	47,585-57,545	30.6	PMEL	9,962	-	-	-	-
85709	KN633923.1	57,546-66,662	29.1	PMEL-like	9,118	-	-	-	-
116317	KN637802.1	5,773-12,720	32.8	B-type SVMP	6,949	22%	3e-85	87.54%	MT032003.1
116318	KN637802.1	12,721-21,599	27.5	SVMP	8,880	19%	0	79.87%	MT070611.1
116319	KN637802.1	21,600-30,546	33.3	B-type SVMP	8,948	3%	1e-53	95.77%	MT032003.1
116320	KN637802.1	30,547-39,768	32.7	A-type SVMP	9,223	13%	0	85.14%	MT124580.1

Table 3.2.1: Details of the 15 outliers extracted from the sliding window analysis. The three first columns include details on the outlier region from the analysis and plot (Figure 3.2.1). The last six columns were drawn directly from the BLAST search of each region, excluding genes within the annotation. Genes shaded in gray were identified using the BLAST, while the others were identified from the *V. berus* gene annotation.

Eurasian populations

The pairwise F_{ST} analysis for the 11 Eurasian populations reflects the differentiation observed in the *V. berus* PCA (Figure 3.1.2), with a range from 0.15 to 0.80 and an average of 0.48. The highest F_{ST} scores were between the remote Sakhalin island and the Western subclade (Switzerland and France) with the most differentiated populations being the Sakhalin island and Northern France (global F_{ST} = 0.80; Figure 3.2.2c). This is as predicted considering the isolation of the Sakhalin population, the presence of *V. b. sachalinensis* in the Far-east, and severe fragmentation within the Western subclade. The geographic distances between the Sakhalin population and Central France is also the furthest in the geographic distance matrix at 9,022 km. The only population that was not extremely differentiated from the Sakhalin population was from Ukraine (global F_{ST} = 0.42).

This polarization was also apparent in the isolation by distance plots (Figure 3.2.2a,b), where the most geographically distant points involve Russia, with the most genetically distant populations being from the Western subclade, and the least from Ukraine. Despite this, the results of the mantel test revealed a positive correlation between geographic (km) and genetic ($F_{ST}/(1 - F_{ST})$) distances (Mantel test: $r = 0.67$, $p < 0.0001$), confirming that there is an isolation by distance effect in Eurasian *V. berus* populations.

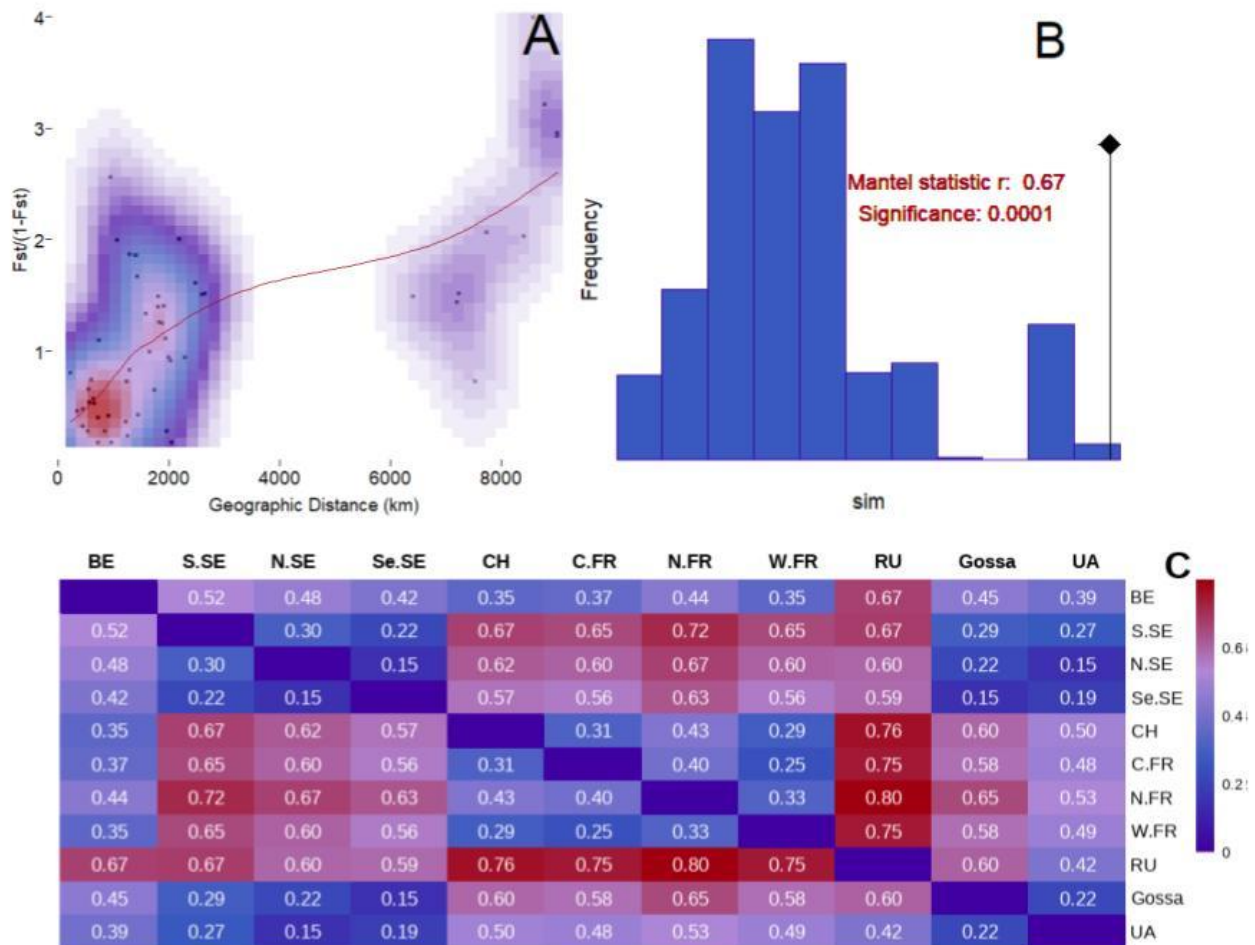


Figure 3.2.2: Genetic differentiation and isolation by distance in *Vipera berus* populations. BE = Belgium, S.SE = Southern Sweden, N.SE = Northern Sweden, Se.SE = South-east Sweden, CH = Switzerland, C.FR = Central France, N.FR = Northern France, W.FR = Western France, RU = Russia, UA = Ukraine. **A)** The regression of the weighted F_{ST} values and geographic distances between populations with $15 > n > 5$. $F_{ST}/(1 - F_{ST})$, recommended by (Rousset 1997) for isolation by distance analyses, was used as the estimate of genetic distance. Euclidean geographic distances between populations are represented on the x-axis. The points in the top right are pairwise comparisons involving the Sakhalin population in the Far East. The red line represents the smoothed local mean, which illustrates a non-linear monotonic relationship of isolation by distance. **B)** A Mantel test using the Spearman method and 9,999 random permutations was performed, resulting in a Mantel statistic of 0.67 and a significance level of 0.0001. The significance of isolation by distance is also demonstrated in the plot. The correlation is represented by the dot on the right, while the histogram distribution displays the nonparametric, permuted data. The isolation of the dot from the histogram exhibits significant spatial structure across Eurasia. **C)** A heatmap of the pairwise weighted F_{ST} values based on the site frequency spectra and genotype likelihood scores. Highly differentiated populations are depicted in red, such as the comparisons between Russia and France. Southern Sweden and Northern France were also heavily differentiated with a pairwise F_{ST} of 0.72. Populations with little divergence, such as the three populations in Sweden and similarly with the French and Swiss populations, are depicted in blue.

V. b. berus vs. *V. b. bosniensis*

The weighted pairwise F_{ST} calculated for *V. b. berus* and *V. b. bosniensis* was much lower ($F_{ST} = 0.207$) than the F_{ST} values calculated between Eurasian populations of *V. berus* (Figure 3.2.2c). However, it reflects the structure of the PCA (Figure 3.1.1) and the sliding window analysis (Figure 3.2.3). The sliding window analysis between *V. b. berus* and *V. b. bosniensis* contained 123,569 windows with 124 windows exceeding the 99.9% quantile threshold for z-scores. To identify the most divergent windows in this quantile, only the windows with z-scores > 5.5 were selected for gene identification, which resulted in 24 remaining windows. Outlier regions were characterized using both the *V. berus* genome annotation and BLAST (Table 3.2.2). The five genes identified using the annotation were: RNase H (Ribonuclease H), KIAA1109, VTA1 (Vesicle Trafficking 1), RAB3Gap2 (RAB3 GTPase activating non-catalytic protein subunit 2), and a BET1-like gene. Nine additional genes were estimated via BLAST, which included DCAF13 (DDB1 and CUL4 associated factor 13), ZNF665-like (Zinc finger protein 665-like), SVMP (Snake Venom Metalloproteinase), AGAP3 (ArfGAP with GTPase domain, ankyrin repeat and PH domain 3), ZNF850-like (Zinc finger protein 850-like), ATF2 (Activating Transcription Factor 2), PPCL (Phosphopantothenate-Cysteine Ligase), PPIH (Peptidyl-Prolyl cis-trans Isomerase H), and ACR-like (Acrosin-like). Windows 520, 30197, 34470, 63048, and 116421 contained no genes identifiable with the annotation or BLAST searching.

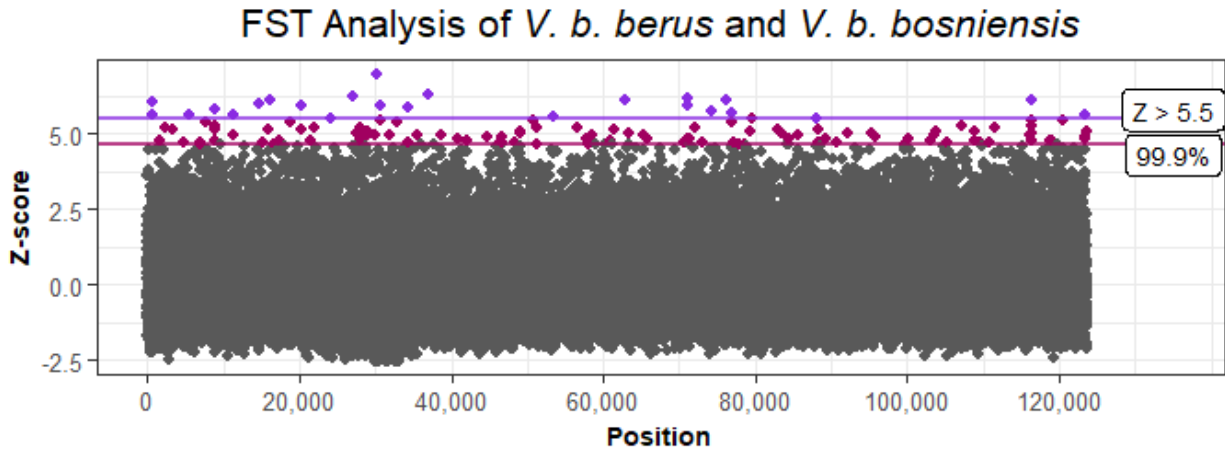


Figure 3.2.3: The plot of a pairwise F_{ST} sliding-window analysis for the Gossa population using 10 kbp windows and steps of 10 kbp. The red line at $y=4.66$ represents the threshold of the 99.9% quantile. Outliers with z-scores > 5.5 are colored in purple as a subset of this quantile to identify regions of interest with noticeable peaks. The plotted dots are the midpoints of each window.

Outlier Region				Gene Identification					
Window number	Scaffold	Position	Z score	Gene(s)	Sites	Query coverage	E-Value	Identity	Accession
517	JTGP01237737.1	8,822-9,661	5.7	DCAF13	841	42%	6e-38	75.88%	XM_015821995.1
520	JTGP01237930.1	9,227-9,910	6.1	-	685	-	-	-	-
5,516	KN614328.1	113,479-123,144	5.7	RNase H	9,667	-	-	-	-
8,811	KN615072.1	138,401-144,893	5.8	BET1-like	6,494	-	-	-	-
11,294	KN615647.1	340,691-348,219	5.6	ZNF665-LIKE	7,530	11%	1e-82	74.51%	XM_026688293.1
14,588	KN616515.1	4,340-7,540	6.0	SVMP	3,202	36%	1e-133	75.39%	MF974507.1
16,057	KN616941.1	78,634-81,634	6.1	SVMP	3,002	56%	0.0	87.50%	MT019621.1
20,354	KN618292.1	31,163-33,249	6.0	AGAP3	2,088	8%	2e-56	90.45%	XM_033166306.1
24,079	KN619498.1	103,186-108,386	5.5	KIAA1109	5,202	-	-	-	-
27,151	KN620478.1	105,869-114,695	6.2	B-type SVMP	8,828	5%	3e-69	93.16%	MT032003.1
30,197	KN621488.1	4,877-5,636	7.0	-	761	-	-	-	-
30,797	KN621701.1	6,737-7,291	6.0	ZNF850-LIKE	556	26%	2e-11	77.70%	XM_032211580.1
34,470	KN623140.1	3,872-4,011	5.9	-	141	-	-	-	-
37,074	KN624312.1	9,049-16,489	6.3	B-type SVMP	7,442	14%	8e-85	84.94%	MT032003.1
53,582	KN631336.1	68,725-72,479	5.6	ATF2	3,756	9%	5e-69	83.02%	KU866087.1
63,048	KN631996.1	12,583-18,403	6.1	-	5,822	-	-	-	-
71,219	KN632629.1	391,873-394,537	6.2	PPCL	2,666	76%	0.0	86.16%	KX211993.1
71,220	KN632629.1	394,538-399,674	5.9	PPIH, PPCL	5,138	19%	4e-141	91.58%	XM_039367437.1
74,369	KN632896.1	209,664-213,889	5.8	ACROSIN-LIKE	4,227	19%	7e-153	93.78%	XM_029286285.1

Outlier Region				Gene Identification					
Window number	Scaffold	Position	Z score	Gene(s)	Sites	Query coverage	E-Value	Identity	Accession
76,286	KN633067.1	52,810-62,208	6.1	VTA1	9,400	-	-	-	-
76,957	KN633124.1	15,180-18,892	5.7	B-type SVMP	3,714	18%	2e-26	82.25%	MT032003.1
88,216	KN634198.1	46,776-53,714	5.5	Rab3gap2	6,940	-	-	-	-
116,421	KN637861.1	24,530-25,300	6.1	-	772	-	-	-	-
123,499	KN639131.1	20,440-30,096	5.6	ATF2	9,658	5%	3e-35	80.70%	KU866087.1

Table 3.2.2: Details of the 24 outliers extracted from the sliding window analysis between *V. b. berus* and *V. b. bosniensis*. The three first columns are drawn from the sliding window analysis (Figure 3.2.1). The last six columns list the details of the BLAST results from searching each region, excluding genes on the annotation. Genes shaded in gray were identified using the BLAST, while white-shaded genes were identified from the *V. berus* gene annotation. Regions that could not be identified with either method are left blank.

3.3 Genome-wide association study (GWAS)

The dataset initially contained 1,778,982 SNPs after repeats were removed, but this was reduced to 370,378 after visualizing the LD decay plot and determining a 0.4 independent pairwise cutoff. After running a PCA, the top four eigenvectors were displayed to visualize the population structure within Gossa (Figure 3.3.1). Although these eigenvectors are low (< 0.1), there are signs of clustering along the first and second principal components. The most notable clusters are on opposite poles of PC1, though this variable does not separate phenotypes.

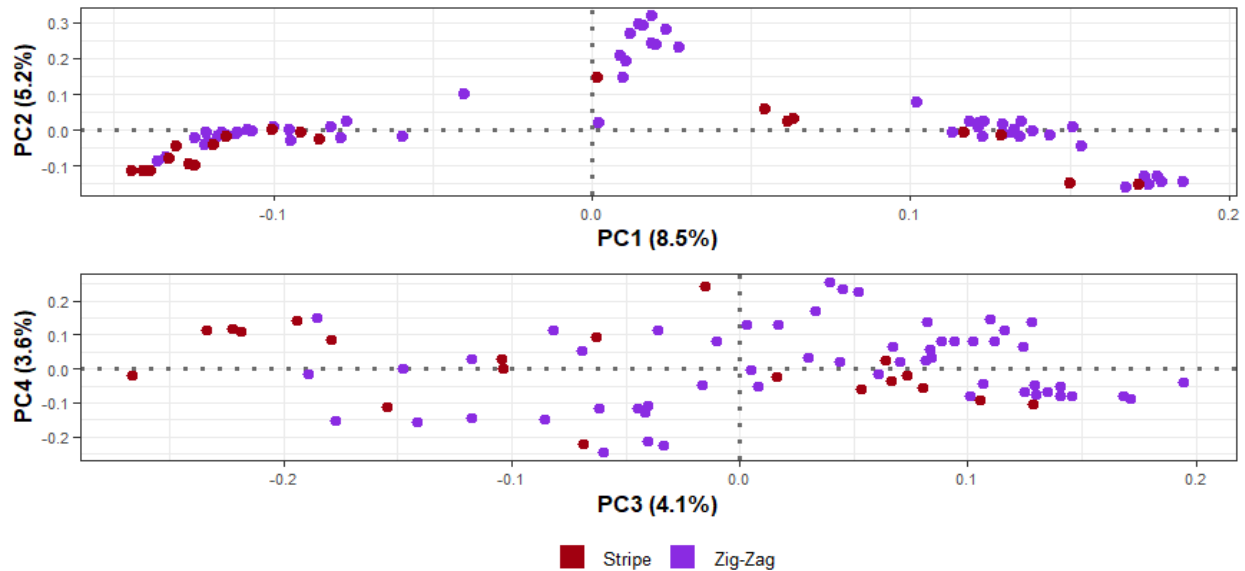


Figure 3.3.1: Population structure on the island of Gossa ($n=77$). The first four variables are all < 0.1 with the highest being 8.5 and the combined weight of 21.4.

A total of 3.5 million SNPs were included in the GWAS. To explore the fit of the data to the model, a quantile-quantile plot was produced to reveal genomic inflation of $\lambda = 1.55$ with the distribution of observed P-values heavily deviating from the null distribution (Figure 3.3.2). Although the distribution of markers displays association along all points, six SNPs were significantly associated ($-\log_{10}P\text{-value} = 5.28$) with pattern variation among the Gossa population. The significant SNPs were identified as KN618963-18581, KN620850-251606, KN633028-20441, KN635276-41139, KN638132-192895 and KN638555-37521 (Table 3.3.1). Of these, only one fell within a gene in the genome annotation: KN638555-37521. The gene in which this SNP is located is UBL5/Hub1 (Ubiquitin-Like 5), which encodes for a 73-amino-acid protein belonging to the UBL family (Chanarat 2021). The only SNP which exceeded the Bonferroni threshold was KN635276-41139, but it did not fall within a gene in the annotation.

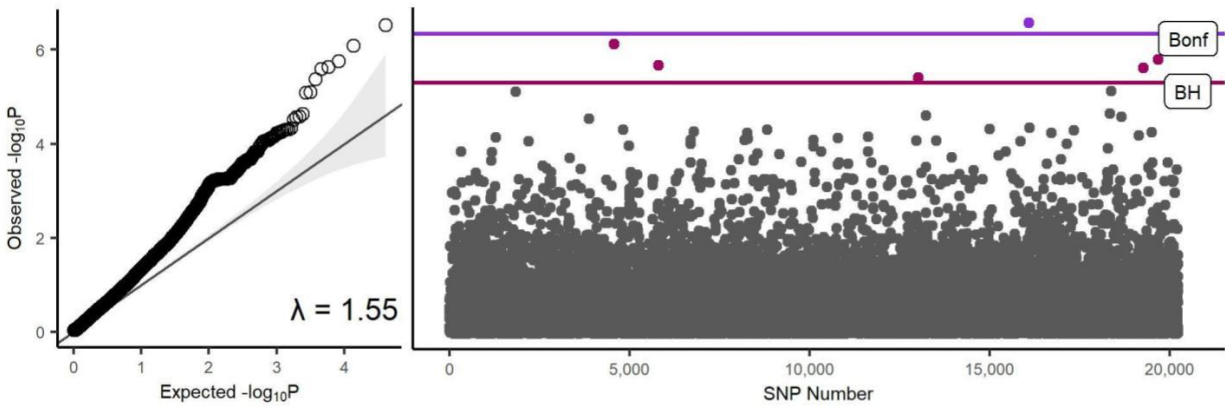


Figure 3.3.2: Quantile-quantile plot and Manhattan plot of P-values from the genome-wide association study between pattern variation in the Gossa population. The thin black line in the quantile-quantile plot signifies the null distribution while the thick black line is composed of each plotted association (20,229). Additionally, the inflation coefficient ($\lambda = 1.55$) is shown in the lower right corner. The Manhattan plot displays the $-\log_{10}P$ -value for each variant with the significance threshold of $-\log_{10}P$ -value = 5.28, as determined by the Benjamini-Hochberg method. The significance line and points exceeding this threshold are marked in red. The purple line represents the Bonferroni threshold ($-\log_{10}P$ -value = 6.31) with one point, marked in purple, exceeding this significance level.

	KN635276	KN618963	KN638555	KN620850	KN638132	KN633028
	-41139	-18581	-37521	-251606	-192895	-20441
Gene	-	-	UBL5/Hub1	-	-	-
P-value	3.00×10^{-7}	8.18×10^{-7}	1.72×10^{-6}	2.29×10^{-6}	2.57×10^{-6}	4.17×10^{-6}

Table 3.3.1: Markers significantly associated with pattern variation on Gossa. The P-value has been drawn from the genome-wide association study and the gene has been identified using the genome annotation. Markers without an identified gene fell in a region on the genome annotation that was not associated with a gene.

Discussion

Reptile pattern polymorphism and the mechanisms driving pattern variation is a vastly under-studied topic, as the majority of pigmentation research is limited to genes specific to melanocytes (KURIYAMA et al. 2020; Rosenblum et al. 2004; Corso et al. 2012; Cox et al. 2013). Pattern variation within snakes plays an enormous role in their ecological standing and can indicate complex interactions which are otherwise difficult to detect (Roulin 2014; MURRAY 2011; Murray & Myerscough 1991; Mckinnon & Pierotti 2010). With the pattern variation of adders in the Gossa population being the only documented occurrence to this degree (Hills et al. 2020) in their entire range, determining the genomic basis is incredibly significant, not only for the Gossa population, but for all wild populations of reptiles that exhibit color/pattern polymorphism. Additionally, the rarity of this occurrence

within their range in comparison to closely related species with smaller ranges led to the assessment of populations across Eurasia in an attempt to determine what enabled *V. b. berus* to expand its range with such success. Unlike previous methods that targeted known pigmentation genes, we took an unbiased approach by scanning the whole genome for associations. In this technique, results lay heavily on the quality of annotation, which ultimately prevented the identification of >70% of candidate regions and five significant SNPs using the genome annotation. Although this figure was reduced to 15% by identifying regions with the BLAST, the percent identity of these results were highly variable and should only be considered potential candidate genes.

Population structure and phylogeny

The phylogenetic relationships inferred here traced the history and migration events of *Vipera* species: *V. berus*, *V. kaznakovi*, *V. renardi*, *V. anatolica*, *V. graeca* and *V. ursinii*. Previous molecular and morphological phylogenetic studies have yielded contradictory results for the relationships within Viperinae while inferring phylogenies based on mtDNA, nDNA and morphological characteristics (Alencar et al. 2016; Freitas et al. 2020; Mizsei et al. 2017; Ferchaud et al. 2012; Zinenko et al. 2015), but there were consistencies in the divergence of *kaznakovi* from the *ursinii-renardi* group (Alencar et al. 2016; Zaher et al. 2019; Figueroa et al. 2016). The placement of *V. anatolica* in this tree is less consistent with recent analyses of mitochondrial genes (Cytochrome b, 12S, and 16S RNA) which estimate a basal divergence of *anatolica* from the *berus, kaznakovi and ursinii* group (Kalyabina-Hauf et al. 2004). Previously however, *anatolica* had initially been considered a species within the *ursinii-renardi* complex, sharing most genetic similarity to *eriwanensis* and *renardi* (Nilson and Andr en 2001), which is what is presented here. The placement of *graeca* under the *renardi* complex has not been observed in any of the phylogenetic analyses of *Vipera* as previously to the work of (Ferchaud et al. 2012), *graeca* had been considered a subspecies of *ursinii*. The split between *ursinii* and the *renardi* complex is estimated to be between 2.4 and 2.0 Ma (Ferchaud et al. 2012), making the relationship between *graeca* and *renardi* unusual.

Within the analysis of *V. berus*, the data confirms a separation of the Northern clade, similar to the results seen in (Ursenbacher et al. 2006a). The results here differ with the division of the majority of Scandinavian populations from the remaining Central European groups with the only individuals not included in the Scandinavian branch being the three Oslo samples, suggesting the adders in southern Norway were last to separate from the Central-Western group, likely due to a second migration event occurring after the last glaciations (Ursenbacher et al. 2006a). The results here also differ with the stark distinction between the Sakhalin population and the remaining ingroup. In this analysis, the Sakhalin population is the basal lineage to all other populations, including the Carpathian group, which supports the subspecies of *V. sachalinensis*. The only population displaying little differentiation between both Sakhalin and the remaining ingroup is the Ukrainian population (Figure 3.1.2). The F_{ST} analysis similarly reflected this with the highest pairwise F_{ST} values involving the Sakhalin population, the lowest of which was between Ukraine and Sakhalin ($F_{ST} = 0.42$; Figure 3.2.2). The Ukrainian population also displayed relatively little variation between the Scandinavian populations, with lower F_{ST} values than Scandinavia had between the remaining Central-European subclade proposed by (Ursenbacher et al. 2006a).

The history of colonization in Europe consists of multiple migration events influenced by stadial and interstadial periods. Adders retreated to refugia accordingly in search of optimal habitat, which created genetic division in the Northern clade. Isolation of distant populations such as *V. sachalinensis* aided in the genetic division between groups, preventing gene flow. Our results reflect this added complexity with low genetic differentiation between Ukraine and Scandinavia suggesting an initial migration event radiating from the Carpathian region into Scandinavia followed by recolonization of Southern Norway from the Central European subclade (Ursenbacher et al. 2006a). This is consistent with the suggestions of a Carpathian refugia by Ursenbacher et al. 2006, as well with other species occupying Europe at this time, such as the field voles (Ursenbacher et al. 2006a; Jaarola & Searle 2002), frogs (Babik et al 2004) and newts (Wielstra et al 2017). In the radiation events following the initial colonization, groups migrated from three additional suggested refugia: East of the Carpathians, Southern France and a centrally located European refugia near the Tatra Mountains in Poland (Ursenbacher et al. 2006a). This recolonization event is likely why the Oslo population displayed less genetic differentiation between the remaining central European group in comparison to the other Scandinavian populations. However, further sampling in southern Norway is necessary to confirm this due to the limited number of samples representing this group in our study.

The long route of migration in addition to the disjunction between the ingroup clearly influenced the correlation between genetic and geographic distances. Although there was a monotonic relationship, genetic and geographic distances were still significantly supported in the mantel test ($R = 0.67$, $p = 0.0001$), proving *V. berus* exhibits isolation by distance across Eurasian populations.

Success of V. berus berus

In the population genetic analysis of *V. b. berus* and *V. b. bosniensis*, 24 highly differentiated candidate regions were discovered through F_{ST} scanning. Within these regions, five candidate genes were identified on the annotation and nine potential candidate genes using the BLAST. The most significant of the genes found was snake venom metalloproteinase (SVMP) Type B, which has been associated with snakes in colder climates. Type B venom is characterized by having hemorrhagic qualities due to the lack of neurotoxic phospholipase A2 (PLA2) in favor of higher snake venom metalloproteinases (SVMPs). Conversely, Type A lacks SVMPs in favor of highly potent neurotoxic venom (Strickland et al. 2018a). Previous studies on *V. berus* venom concluded that neurotoxic symptoms have been observed and proven in populations of *V. b. bosniensis*, whereas the occurrences of neurotoxic symptoms in the nominate subspecies are incredibly rare, only being documented in localized Hungarian populations (Magdalan et al. 2010; Malina et al. 2008; Malina et al. 2013). As with pattern variation in snakes, venom composition is highly informative of their feeding and defense strategies, in addition to their fitness (Strickland et al. 2018b). An analysis of venom variation among Mojave rattlesnakes (*Crotalus scutulatus*) concluded that venom phenotype is correlated with climatic variables. In the case of *Crotalus scutulatus*, minimum annual temperatures and minimum recorded temperature were significantly associated with the presence of Type B venom. As many factors contribute to venom composition, there is no definitive answer to why there is this climatic association, but there is strong evidence that SVMPs contribute to the efficiency of digestion in colder

conditions (Thomas & Pough 1979), possibly enabling snakes with highly proteolytic venom to expand their range to more diverse climates as well as expand their diet to encompass variably sized prey (Strickland et al. 2018b; Mackessy 2010). If this is the case, it is possible that SVMs enabled *V. b. berus* to successfully colonize the range they did because it allowed for them to survive variable climatic conditions with diverse prey.

Gossa pattern variation

Genome-wide association studies (GWAS) are popular and sensitive analyses that are used to identify significant SNPs across the genome that share association with a trait. Most often, GWAS require large sample sizes to increase the power of the association, especially in studies that contain highly related individuals, such as the Gossa population. In the preliminary tests of Gossa, two pregnant striped vipers were captured to study the inheritance of the striped phenotype. Both mothers and all neonates (n=16, 7 striped, 9 non striped) were sampled prior to releasing them. With the collection of snakes being limited to one season and the rarity of striped individuals, it was decided that these samples would be included in the analysis. To counter the chance of genomic inflation and population structure, the first six principal components were used as covariates in the GWAS. The result still displayed some inflation ($\lambda = 1.55$), however six significant SNPs were still identified using the Benjamini-Hochberg method. Only one SNP was identified using the annotation as the rest did not intersect with a gene. The SNP identified was UBL5/Hub1, a ubiquitin-like protein

In the F_{ST} analysis between striped and unstriped individuals, 15 candidate regions were detected. Several notable hits such as PMEL, SUOX, DGKA, DDN and SVM type-A were discovered, four of which were identified using the genome annotation (Table 3.2.1). Previous research on SUOX, DGKA and PMEL mutations outline various phenotypes in vertebrates, from arachnomelia in cattle, sulfocysteinuria in humans, and as earlier mentioned, pigmentation disorders through the modification of PMEL fibrils across species (Claerhout et al. 2018; Qi et al 2018; Drögemüller et al. 2010; Watt et al. 2011). The SUOX (sulfite oxidase) gene encodes for an enzyme which is a catalyst to the oxidation of sulfite. SUOX mutations are most often associated with sulfite oxidase deficiency (SOD), however it also has a significant role in pigmentation, similarly to PMEL and the DGK family (Qi et al 2018; Xie & Song 2022; Read 2016). Isoforms of the DGK family have been suggested to regulate tyrosinase expression, which in turn impacts melanin production as tyrosinase is a catalyst of melanogenesis (Kawaguchi et al. 2012). The role of RAB5b in pigmentation is primarily in the uptake of melanosomes by the keratinocytes. RAB5b regulates the early stages of the endocytic pathway and has been observed surrounding melanosomes (Correia et al. 2018).

The result of DDN from the BLAST is interesting because it has not previously been associated with pigmentation studies. Little is known about DDN (Dendrin) other than its expression in the brain, its role in synaptic plasticity and the novel discovery in the kidneys (Herb et al. 1997; Ji et al. 2019). It is believed to be involved in postsynaptic cytoskeletal modulation, where it interacts with α -actinin in neural dendrites. Although the formation of dendrites in melanocytes is somewhat inconclusive, parallels between neural cells and melanocyte morphology have led to the discovery that similar mechanisms take place in the dendrite formation of both (Glynis 2002; Mengnan et al. 2020). With the function of dendrin

still largely unknown, it may be that dendrin has a similar role in melanocytes, leading to pattern formation. In *Phelsuma* geckos, the occurrence of spots and stripes were found to be caused by melanophores stretching their dendrites through the layers of iridophores and depositing melanin granules above them in the dermis (Saenko et al. 2013). Although both dermal and epidermal pigmentation contribute to the pattern formation in reptiles (Alibardi 2013), this study suggested melanin granules contribute primarily for pattern morphology while darker and larger melanophores are suggested to counter UV radiation, which emphasizes the importance of dendrite formation in pattern variation of reptiles.

The morphological variation observed on the island of Gossa is exceptionally rare, being the only recorded occurrence to this degree in their entire range. This striped phenotype is not however unique to only one taxa in the genus of *Vipera*. Similar polymorphism occurs in *Vipera aspis* (MRCA = 15 Ma) and *Vipera seoanei* (MRCA = 4 Ma; Šmíd & Tolley 2019), which are suggested to be environmentally-driven as the melanistic phenotype is (Martínez-Freiría & Brito 2013; Martínez-Freiría et al. 2020). There are four phenotypes across the range of *V. aspis*: striped, zig-zag, melanistic and concolor (Ducrest et al. 2014). These phenotypes are also observed across the range of *V. seoanei* in addition to the cantabrica pattern which is primarily observed in the Cantabrian region (Lucchini et al. 2020). Although the most common phenotype within the range of *V. berus* is the zig-zag, few localized populations have displayed high proportions of melanistic individuals in high altitudes (Cui et al. 2016). Melanism in *Vipera* has been proven in *V. seoanei* to follow the “thermal melanism hypothesis” in which the dark pigment is used to heighten the efficiency of thermoregulation, enabling individuals to hunt for longer periods and reduce time required for basking, driving selection. Although the island of Gossa falls into the categories of high altitude and high latitude, there have been no observations of melanistic individuals there. In the analysis of *V. seoanei*, there is a clear trade-off for melanism in populations (i.e. longer hunting hours), but the potential benefit of the striped phenotype is less obvious. In many species of snakes, longitudinal stripes occur for a rapid-escape strategy while cryptic patterns such as the zig-zag phenotype occur for aposematism (Wolf & Werner 1994; Martínez-Freiría et al. 2017). Though the striped phenotype is more conspicuous, the benefit of the stripe is primarily in the escape from predation rather than to avoid detection (Martínez-Freiría et al. 2017).

Interestingly, the POMC gene was not correlated with pattern variation in the Gossa population as it was in the closely related asp viper (*Vipera aspis*; Ducrest et al. 2014). In the analysis of color polymorphism in the asp viper, two genes were screened to test for correlation of polymorphism: POMC and MC1R. Both genes are related to melanin specific pigmentation, although POMC also contributes to several other functions and hormones, making it highly conserved across vertebrates (Cai & Hruby 2016). Instead, PMEL, an important component of fibrillation in pre-melanosomes, was identified as being significantly associated with polymorphism on Gossa.

PMEL (Pre-melanosomal protein) plays a vital role in the development of melanosomes by creating tight parallel sheets of fibrils for melanin to synthesize and adhere to in stage III and IV of melanosome biogenesis (Bissig et al. 2016). These tight fibrils give the melanosomes their ellipsoidal shape, characteristic of the darkly pigmented eumelanin. Pheomelanin-producing melanocytes (red and yellow pigment) lack this ellipsoidal shape and are instead spherical, similarly to their shape in PMEL knock-out experiments (Hellström et al. 2011). PMEL is very tightly regulated in the production of melanosomes by several

subdomains which ensure correct trafficking due to the danger of amyloid structures (Bissig et al. 2016). Mutations in PMEL that impact this tightly regulated trafficking ultimately cause defects in fibrillogenesis of the stage I and II melanosomes. Mutations that impact PMEL oligomerization are particularly dangerous as this converts the amyloidogenesis into a pathogenic process (Bissig et al. 2016). Therefore, mutations in PMEL have possibly not been observed in wild populations of snakes because of its high risk.

Although these closely related snakes have all evolved the striped and melanistic phenotypes in addition to the cryptic phenotype, they differ in the specific genetic mechanisms used to achieve this variation. The advantages of species to adapt color polymorphism according to their environment is highly beneficial in the vast range that European vipers inhabit, and as a consequence, these vipers have evolved similar phenotypes in parallel (Martínez-Freiría et al. 2017; Lucchini et al. 2020; Martínez-Freiría & Brito 2013; Martínez-Freiría et al. 2020).

References

- Alibardi, L., 2013. Observations on the ultrastructure and distribution of chromatophores in the skin of chelonians. *Acta Zoologica* 94, 222–232.. doi:10.1111/j.1463-6395.2011.00546.x
- Allen, W.L., Baddeley, R., Scott-Samuel, N.E., Cuthill, I.C., 2013. The evolution and function of pattern diversity in snakes. *Behavioral Ecology* 24, 1237–1250.. doi:10.1093/beheco/art058
- Andersson S, Johansson L. Cold hardiness in the boreal adder, *Vipera berus*. *Cryo Letters*. 2001 May-Jun;22(3):151-156. PMID: 11788854.
- Andersson, S., 2003. Hibernation, habitat and seasonal activity in the adder, *Vipera berus*, north of the Arctic Circle in Sweden. *Amphibia-Reptilia*, 24(4), pp.449-457.
- Andrén, C., Nilson, G., 1981. Reproductive success and risk of predation in normal and melanistic colour morphs of the adder, *Vipera berus*. *Biological Journal of the Linnean Society* 15, 235–246.. doi:10.1111/j.1095-8312.1981.tb00761.x
- Babik, W., Branicki, W., Sandera, M., Litvinchuk, S., Borkin, L.J., Irwin, J.T., Rafiński, J., 2004. Mitochondrial phylogeography of the moor frog, *Rana arvalis*. *Molecular Ecology* 13, 1469–1480.. doi:10.1111/j.1365-294x.2004.02157.x
- Batai, K., Cui, Z., Arora, A., Shah-Williams, E., Hernandez, W., Ruden, M., Hollowell, C.M.P., Hooker, S.E., Bathina, M., Murphy, A.B., Bonilla, C., Kittles, R.A., 2021. Genetic loci associated with skin pigmentation in African Americans and their effects on vitamin D deficiency. *PLOS Genetics* 17, e1009319.. doi:10.1371/journal.pgen.1009319
- Bauwens, D., Claus, K., 2019. Seasonal variation of mortality, detectability, and body condition in a population of the adder (*Vipera berus*). *Ecology and Evolution* 9, 5821–5834.. doi:10.1002/ece3.5166
- Beissinger, T. M., HIRSCH, C. N., VAILLANCOURT, B., DESHPANDE, S., BARRY, K., BUELL, C. R., KAEPLER, S. M., GIANOLA, D. & DE LEON, N. 2014. A Genome-Wide Scan for Evidence of Selection in a Maize Population Under Long-Term Artificial Selection for Ear Number. *Genetics*, 196, 829-840.
- Beissinger, T.M., Rosa, G.J., Kaeppler, S.M., Gianola, D., De Leon, N., 2015. Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genetics Selection Evolution* 47.. doi:10.1186/s12711-015-0105-9
- Bissig, C., Rochin, L., Van Niel, G., 2016. PMEL Amyloid Fibril Formation: The Bright Steps of Pigmentation. *International Journal of Molecular Sciences* 17, 1438.. doi:10.3390/ijms17091438
- Cai, M. and J Hruby, V., 2016. The melanocortin receptor system: a target for multiple degenerative diseases. *Current Protein and Peptide Science*, 17(5), pp.488-496.
- Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., Yandell, M., 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* 18, 188–196.. doi:10.1101/gr.6743907
- Carson, A.R., Smith, E.N., Matsui, H., Brækkan, S.K., Jepsen, K., Hansen, J.-B., Frazer, K.A., 2014. Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics* 15, 125.. doi:10.1186/1471-2105-15-125

- Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S.T., Sinding, M.H.S., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., Gilbert, M.T.P., 2018. Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution* 9, 410–419.. doi:10.1111/2041-210x.12871
- CHANARAT, S. 2021. UBL5/Hub1: An Atypical Ubiquitin-Like Protein with a Typical Role as a Stress-Responsive Regulator. *International Journal of Molecular Sciences*, 22, 9384.
- Chang, C., Wu, P., Baker, R.E., Maini, P.K., Alibardi, L., Chuong, C.-M., 2009. Reptile scale paradigm: Evo-Devo, pattern formation and regeneration. *The International Journal of Developmental Biology* 53, 813–826.. doi:10.1387/ijdb.072556cc
- Christoforou, A., Dondrup, M., Mattingsdal, M., Mattheisen, M., Giddaluru, S., Markus, Rietschel, M., Cichon, S., Djurovic, S., Ole, Jonassen, I., Vidar, Puntervoll, P., Stéphanie, 2012. Linkage-Disequilibrium-Based Binning Affects the Interpretation of GWASs. *The American Journal of Human Genetics* 90, 727–733.. doi:10.1016/j.ajhg.2012.02.025
- Claerhout, H., Witters, P., Régál, L., Jansen, K., Van Hoestenbergh, M.-R., Breckpot, J., Vermeersch, P., 2018. Isolated sulfite oxidase deficiency. *Journal of Inherited Metabolic Disease* 41, 101–108.. doi:10.1007/s10545-017-0089-4
- Correia, M.S., Moreiras, H., Pereira, F.J.C., Neto, M.V., Festas, T.C., Tarafder, A.K., Ramalho, J.S., Seabra, M.C., Barral, D.C., 2018. Melanin Transferred to Keratinocytes Resides in Nondegradative Endocytic Compartments. *Journal of Investigative Dermatology* 138, 637–646.. doi:10.1016/j.jid.2017.09.042
- Corso, J., Gonçalves, G.L. and de Freitas, T.R., 2012. Sequence variation in the melanocortin-1 receptor (MC1R) pigmentation gene and its role in the cryptic coloration of two South American sand lizards. *Genetics and Molecular Biology*, 35(1), pp.81-87.
- Cox, C.L., Rabosky, A.R.D. & Chippindale, P.T., 2013. Sequence variation in the Mc1r gene for a group of polymorphic snakes. *Gene*, 513(2), pp.282–286.
- Cui, S., Luo, X., Chen, D., Sun, J., Chu, H., Li, C., Jiang, Z., 2016. The adder (*Vipera berus*) in Southern Altay Mountains: population characteristics, distribution, morphology and phylogenetic position. *PeerJ* 4, e2342.. doi:10.7717/peerj.2342
- Dean, J., Ghemawat, S., 2008. MapReduce. *Communications of the ACM* 51, 107–113.. doi:10.1145/1327452.1327492
- Dores, R.M., Liang, L., Davis, P., Thomas, A.L., Petko, B., 2016. 60 YEARS OF POMC: Melanocortin receptors: evolution of ligand selectivity for melanocortin peptides. *Journal of Molecular Endocrinology* 56, T119–T133.. doi:10.1530/jme-15-0292
- Drögemüller, C., Tetens, J., Sigurdsson, S., Gentile, A., Testoni, S., Lindblad-Toh, K., Leeb, T., 2010. Identification of the Bovine Arachnomelia Mutation by Massively Parallel Sequencing Implicates Sulfite Oxidase (SUOX) in Bone Development. *PLOS Genetics* 6, e1001079.. doi:10.1371/journal.pgen.1001079
- Ducrest, A.-L., Ursenbacher, S., Golay, P., Monney, J.-C., Mebert, K., Roulin, A., Dubey, S., 2014. Pro-opiomelanocortin gene and melanin-based colour polymorphism in a reptile. *Biological Journal of the Linnean Society* 111, 160–168.. doi:10.1111/bij.12182
- Fadista, J., Manning, A. K., Florez, J. C., & Groop, L. (2016). The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European journal of human genetics : EJHG*, 24(8), 1202–1205. <https://doi.org/10.1038/ejhg.2015.269>

- Glynis Scott (2002). Rac and Rho: The Story Behind Melanocyte Dendrite Formation. , 15(5), 322–330. doi:10.1034/j.1600-0749.2002.02056.x
- Gur, D., Bain, E.J., Johnson, K.R., Aman, A.J., Pasoili, H.A., Flynn, J.D., Allen, M.C., Deheyn, D.D., Lee, J.C., Lippincott-Schwartz, J., Parichy, D.M., 2020. In situ differentiation of iridophore crystallotypes underlies zebrafish stripe patterning. *Nature Communications* 11.. doi:10.1038/s41467-020-20088-1
- Hahn, M. W. (2018). *Molecular Population Genetics*, Oxford University Press.
- Hamazaki, K., & Iwata, H. (2020). RAINBOW: Haplotype-based genome-wide association study using a novel SNP-set method. *PLoS computational biology*, 16(2), e1007663. <https://doi.org/10.1371/journal.pcbi.1007663>
- Hellström, A.R., Watt, B., Fard, S.S., Tenza, D., Mannström, P., Narfström, K., Ekestén, B., Ito, S., Wakamatsu, K., Larsson, J., Ulfendahl, M., Kullander, K., Raposo, G., Kerje, S., Hallböök, F., Marks, M.S., Andersson, L., 2011. Inactivation of Pmel Alters Melanosome Shape But Has Only a Subtle Effect on Visible Pigmentation. *PLOS Genetics* 7, e1002285.. doi:10.1371/journal.pgen.1002285
- Herb A., Wisden W., Catania M. V., Maréchal D., Dresse A., Seeburg P. H., Prominent Dendritic Localization in Forebrain Neurons of a Novel mRNA and Its Product, Dendrin, *Molecular and Cellular Neuroscience*, Volume 8, Issue 5, 1997, Pages 367-374, ISSN 1044-7431, <https://doi.org/10.1006/mcne.1996.0594>.
- HILLS, L., Lewis, B., Hills, R., 2020. Dorsal stripe polymorphism of *Vipera berus* in south-east England. *The Herpetological Bulletin* 26-28.
- Horreo, J.L., Fitze, P.S., 2018. Postglacial Colonization of Northern Europe by Reptiles, in: . pp. 197–214.. doi:10.1007/978-3-319-95954-2_12
- Jaarola, M., Searle, J.B., 2002. Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology* 11, 2613–2621.. doi:10.1046/j.1365-294x.2002.01639.x
- Ji, Z., Li, H., Yang, Z., Huang, X., Ke, X., Ma, S., Lin, Z., Lu, Y., Zhang, M., 2019. Kibra Modulates Learning and Memory via Binding to Dendrin. *Cell Reports* 26, 2064–2077.e7.. doi:10.1016/j.celrep.2019.01.097
- Kalyabina-Hauf, S., Schweiger, S., Joger, U. and Mayer, W., 2004. Phylogeny and systematics of adders (*Vipera bents* complex).
- Kawaguchi, M., Valencia, J.C., Namiki, T., Suzuki, T., Hearing, V.J., 2012. Diacylglycerol Kinase Regulates Tyrosinase Expression and Function in Human Melanocytes. *Journal of Investigative Dermatology* 132, 2791–2799.. doi:10.1038/jid.2012.261
- Korf I. Gene finding in novel genomes. *BMC Bioinformatics*. 2004 May 14;5:59
- Korneliusson TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356
- Krude, H., Biebermann, H., Luck, W., Horn, R., Brabant, G., Grüters, A., 1998. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nature Genetics* 19, 155–157.. doi:10.1038/509
- KURIYAMA, T., MURAKAMI, A., BRANDLEY, M. & HASEGAWA, M. 2020. Blue, black, and stripes: evolution and development of color production and pattern formation in lizards and snakes. *Frontiers in Ecology and Evolution*, 8, 232.
- Lefort, V., Desper, R., Gascuel, O., 2015. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program: Table 1.. *Molecular Biology and Evolution* 32, 2798–2800.. doi:10.1093/molbev/msv150

- Lourdais, O., Guillon, M., DeNardo, D. and Blouin-Demers, G., 2013. Cold climate specialization: adaptive covariation between metabolic rate and thermoregulation in pregnant vipers. *Physiology & behavior*, 119, pp.149-155.
- Lucchini, N., Kaliontzopoulou, A., Val, G.A., Martínez-Freiría, F., 2020. Sources of intraspecific morphological variation in *Vipera seoanei*: allometry, sex, and colour phenotype. *Amphibia-Reptilia* 42, 1–16.. doi:10.1163/15685381-bja10024
- Magdalan J., Trocha M., Merwid-Ląd A., Sozański T., Zawadzki M., *Vipera berus* Bites in the Region of Southwest Poland—A Clinical Analysis of 26 Cases, *Wilderness & Environmental Medicine*, Volume 21, Issue 2, 2010, Pages 114-119, ISSN 1080-6032, <https://doi.org/10.1016/j.wem.2010.01.005>.
- Malina, T., Babocsay, G., Krecsák, L., Erdész, C., 2013. Further Clinical Evidence for the Existence of Neurotoxicity in a Population of the European Adder (*Vipera berus berus*) in Eastern Hungary: Second Authenticated Case. *Wilderness & Environmental Medicine* 24, 378–383.. doi:10.1016/j.wem.2013.06.005
- Malina, T., Krecsák, L., Jelić, D., Maretić, T., Tóth, T., Šiško, M., Pandak, N., 2011. First clinical experiences about the neurotoxic envenomings inflicted by lowland populations of the Balkan adder, *Vipera berus bosniensis*. *NeuroToxicology* 32, 68–74.. doi:10.1016/j.neuro.2010.11.007
- Malina, T., Krecsak, L., Warrell, D.A., 2008. Neurotoxicity and hypertension following European adder (*Vipera berus berus*) bites in Hungary: case report and review. *QJM: An International Journal of Medicine* 101, 801–806.. doi:10.1093/qjmed/hcn079
- Martin, M. D., Dolmen, D., Ophus, K. H., Bringsøe, H., Nielsen, R., & Allentoft, M. (2018). Biodiversity assessment and genomic analysis of a unique population of *Vipera berus* (Rep.).
- Martínez-Freiría, F., Brito, J.C., 2013. Integrating classical and spatial multivariate analyses for assessing morphological variability in the endemic Iberian viper *Vipera seoanei*. *Journal of Zoological Systematics and Evolutionary Research* 51, 122–131.. doi:10.1111/jzs.12015
- Martínez-Freiría, F., Toyama, K.S., Freitas, I., Kaliontzopoulou, A., 2020. Thermal melanism explains macroevolutionary variation of dorsal pigmentation in Eurasian vipers. *Scientific Reports* 10.. doi:10.1038/s41598-020-72871-1
- Martínez-Freiría, F., Lanuza, G. P., Pimenta, A. A. , Tiago Pinto, Xavier Santos, Aposematism and crypsis are not enough to explain dorsal polymorphism in the Iberian adder, *Acta Oecologica*, Volume 85, 2017, Pages 165-173, ISSN 1146-609X, <https://doi.org/10.1016/j.actao.2017.11.003>.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Mckinnon, J.S., Pierotti, M.E.R., 2010. Colour polymorphism and correlated characters: genetic mechanisms and evolution. *Molecular Ecology* 19, 5101–5125.. doi:10.1111/j.1365-294x.2010.04846.x
- Mengnan Li, Sina K. Knapp, Sandra Iden, Mechanisms of melanocyte polarity and differentiation: What can we learn from other neuroectoderm-derived lineages?, *Current Opinion in Cell Biology*, Volume 67, 2020, Pages 99-108, ISSN 0955-0674, <https://doi.org/10.1016/j.ceb.2020.09.001>.

- MURRAY, J. D. 2011. *Mathematical Biology II: Spatial Models and Biomedical Applications*, Springer New York.
- MURRAY, J. D. & MYERSCOUGH, M. R. 1991. Pigmentation pattern formation on snakes. *J Theor Biol*, 149, 339-360.
- Niskanen, M. and Mappes, J., 2005. Significance of the dorsal zigzag pattern of *Vipera latastei gaditana* against avian predators. *Journal of animal ecology*, 74(6), pp.1091-1101.
- Nilson, G., 1980. Male Reproductive Cycle of the European Adder, *Vipera berus*, and Its Relation to Annual Activity Periods. *Copeia* 1980, 729.. doi:10.2307/1444451
- Nilson, G. and Andrén, C., 2001. The meadow and steppe vipers of Europe and Asia—the *Vipera* (Acridophaga) *ursinii* complex. *Acta Zoologica Academiae Scientiarum Hungaricae*, 47(2-3), pp.87-267.
- Nilson, G., ANDRÉN, C. & SZYNDLAR, Z. 1994. The systematic position of the Common Adder, *Vipera berus* (L.) (Reptilia, Viperidae), in North Korea and adjacent regions. *Bonner zoologische Beiträge : Herausgeber: Zoologisches Forschungsinstitut und Museum Alexander Koenig, Bonn*, 45, 49-56.
- OPHUS, K. H. 2017. DNA sampling of the adder *Vipera berus* on the island of Gossen, Aukra municipality, Norway. NTNU Vitenskapsmuseet, The Natural History Museum of Denmark, and The University of California, Berkeley.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.
- QI, Z., XIE, S., CHEN, R., AISA, H. A., HON, G. C. & GUAN, Y. 2018. Tissue-specific Gene Expression Prediction Associates Vitiligo with SUOX through an Active Enhancer. *bioRxiv*, 337196.
- Read, Jazlyn (2016). Familial melanoma risk genes in Queensland. MPhil Thesis, School of Medicine, The University of Queensland. doi: 10.14264/uql.2016.393
- Risch, N.; Merikangas, K. (1996). The Future of Genetic Studies of Complex Human Diseases. , 273(5281), 1516–1517. doi:10.1126/science.273.5281.1516
- Rosenblum, E.B., Hoekstra, H.E., Nachman, M.W., 2004. ADAPTIVE REPTILE COLOR VARIATION AND THE EVOLUTION OF THE MC1R GENE. *Evolution* 58, 1794–1808.. doi:10.1111/j.0014-3820.2004.tb00462.x
- Roulin, A., 2004. The evolution, maintenance and adaptive function of genetic colour polymorphism in birds. *Biological Reviews* 79, 815–848.. doi:10.1017/s1464793104006487
- Rousset, F., 1997. Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. *Genetics* 145, 1219–1228.. doi:10.1093/genetics/145.4.1219
- Saenko, S.V., Teyssier, J., Van Der Marel, D., Milinkovitch, M.C., 2013. Precise colocalization of interacting structural and pigmentary elements generates extensive color pattern variation in Phelsumalizards. *BMC Biology* 11, 105.. doi:10.1186/1741-7007-11-105
- Santos, X., Vidal-García, M., Brito, J.C., Fahd, S., Llorente, G.A., Martínez-Freiría, F., Parellada, X., Pleguezuelos, J.M., Sillero, N., 2014. Phylogeographic and environmental correlates support the cryptic function of the zigzag pattern in a

- European viper. *Evolutionary Ecology* 28, 611–626..
doi:10.1007/s10682-014-9699-6
- Schubert M, Ermini L, Sarkissian CD, Jónsson H, Ginolhac A, Schaefer R, Martin MD, Fernández R, Kircher M, McCue M, Willerslev E, and Orlando L. "Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX". *Nat Protoc.* 2014 May;9(5):1056-82. doi: 10.1038/nprot.2014.063. Epub 2014 Apr 10. PubMed PMID: 24722405
- Smit, A. F. A., Hubler, R., & Green, P., 2015. RepeatMasker Open-4.0. Institute for Systems Biology. <https://www.repeatmasker.org>
- SMITH, M. 1951. *The British amphibians and reptiles*. London.
- Stahl, K., Gola, D. and König, I.R., 2021. Assessment of Imputation Quality: Comparison of Phasing and Imputation Algorithms in Real Data. *Frontiers in genetics*, 12.
- Stamatakis, A., 2016. *The RAxML v8. 2. X Manual*. Heidelberg Institute for Theoretical Studies. Available at: <https://cme.h-its.org/exelixis/resource/download/NewManual.pdf>.
- STANKE, M., KELLER, O., GUNDUZ, I., HAYES, A., WAACK, S. & MORGENSTERN, B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research*, 34, W435-W439.
- Stephen P. Mackessy, evolutionary trends in venom composition in the Western Rattlesnakes (*Crotalus viridis sensu lato*): Toxicity vs. tenderizers, *Toxicon*, Volume 55, Issue 8, 2010, Pages 1463-1474, ISSN 0041-0101, <https://doi.org/10.1016/j.toxicon.2010.02.028>.
- Strickland, Jason & Mason, Andrew & Rokyta, Darin & Parkinson, Christopher. (2018a). Phenotypic Variation in Mojave Rattlesnake (*Crotalus scutulatus*) Venom Is Driven by Four Toxin Families. *Toxins*. 10. 10.3390/toxins10040135.
- Strickland, J.L., Smith, C.F., Mason, A.J., Schield, D.R., Borja, M., Castañeda-Gaytán, G., Spencer, C.L., Smith, L.L., Trápaga, A., Bouzid, N.M., Campillo-García, G., Flores-Villela, O.A., Antonio-Rangel, D., Mackessy, S.P., Castoe, T.A., Rokyta, D.R., Parkinson, C.L., 2018b. Evidence for divergent patterns of local selection driving venom variation in Mojave Rattlesnakes (*Crotalus scutulatus*). *Scientific Reports* 8.. doi:10.1038/s41598-018-35810-9
- Sutton, B.S., Langefeld, C.D., Williams, A.H., Norris, J.M., Saad, M.F., Haffner, S.M. and Bowden, D.W. (2005), Association of Proopiomelanocortin Gene Polymorphisms with Obesity in the IRAS Family Study. *Obesity Research*, 13: 1491-1498. <https://doi.org/10.1038/oby.2005.180>
- Šmíd, J., Tolley, K.A., 2019. Calibrating the tree of vipers under the fossilized birth-death model. *Scientific Reports* 9.. doi:10.1038/s41598-019-41290-2
- Theron, E., Hawkins, K., Bermingham, E., Ricklefs, R.E., Mundy, N.I., 2001. The molecular basis of an avian plumage polymorphism in the wild. *Current Biology* 11, 550–557.. doi:10.1016/s0960-9822(01)00158-0
- Thomas R.G., Pough F.H., The effect of rattlesnake venom on digestion of prey, *Toxicon*, Volume 17, Issue 3, 1979, Pages 221-228, ISSN 0041-0101, [https://doi.org/10.1016/0041-0101\(79\)90211-3](https://doi.org/10.1016/0041-0101(79)90211-3).
- Tzika, A.C., Ullate-Agote, A., Grbic, D., Milinkovitch, M.C., 2015. Reptilian Transcriptomes v2.0: An Extensive Resource for Sauropsida Genomics and Transcriptomics. *Genome Biology and Evolution* 7, 1827–1841.. doi:10.1093/gbe/evv106

- Ullate-Agote, A., Burgelin, I., Debry, A., Langrez, C., Montange, F., Peraldi, R., Daraspe, J., Kaessmann, H., Milinkovitch, M.C. and Tzika, A.C., 2020. Genome mapping of a LYST mutation in corn snakes indicates that vertebrate chromatophore vesicles are lysosome-related organelles. *Proceedings of the National Academy of Sciences*, 117(42), pp.26307-26317.
- Ursenbacher, S., Carlsson, M., Helfer, V., Tegelström, H., Fumagalli, L., 2006a. Phylogeography and Pleistocene refugia of the adder (*Vipera berus*) as inferred from mitochondrial DNA sequence data. *Molecular Ecology* 15, 3425–3437.. doi:10.1111/j.1365-294x.2006.03031.x
- Ursenbacher, S., Conelli, A., Golay, P., Monney, J.C., Zuffi, M.A.L., Thiery, G., Durand, T. and Fumagalli, L., 2006b. Phylogeography of the asp viper (*Vipera aspis*) inferred from mitochondrial DNA sequence data: evidence for multiple Mediterranean refugial areas. *Molecular Phylogenetics and Evolution*, 38(2), pp.546-552.
- Ursenbacher, S., Monney, J.-C., Fumagalli, L., 2009. Limited genetic diversity and high differentiation among the remnant adder (*Vipera berus*) populations in the Swiss and French Jura Mountains. *Conservation Genetics* 10, 303–315.. doi:10.1007/s10592-008-9580-7
- Ursenbacher S., Schweiger S, Tomović L., Crnobrnja-Isailović J., Fumagalli L., Mayer W., Molecular phylogeography of the nose-horned viper (*Vipera ammodytes*, Linnaeus (1758)): Evidence for high genetic diversity and multiple refugia in the Balkan peninsula, *Molecular Phylogenetics and Evolution*, Volume 46, Issue 3, 2008, Pages 1116-1128, ISSN 1055-7903, <https://doi.org/10.1016/j.ympev.2007.11.002>.
- Vieira, F.G., Lassalle, F., Korneliussen, T.S., Fumagalli, M., 2016. Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biological Journal of the Linnean Society* 117, 139–149.. doi:10.1111/bij.12511
- Watt, B., Tenza, D., Lemmon, M.A., Kerje, S., Raposo, G., Andersson, L., Marks, M.S., 2011. Mutations in or near the Transmembrane Domain Alter PMEL Amyloid Formation from Functional to Pathogenic. *PLOS Genetics* 7, e1002286.. doi:10.1371/journal.pgen.1002286
- Westerström, A., Petrov, B. and Tzankov, N., 2010. Envenoming following bites by the Balkan adder *Vipera berus bosniensis*—First documented case series from Bulgaria. *Toxicon*, 56(8), pp.1510-1515.
- Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- Wielstra, B., Zieliński, P., Babik, W., 2017. The Carpathians hosted extra-Mediterranean refugia-within-refugia during the Pleistocene Ice Age: genomic evidence from two newt genera. *Biological Journal of the Linnean Society* 122, 605–613.. doi:10.1093/biolinnean/blx087
- Wolf, M. and Werner, Y.L., 1994. The striped colour pattern and striped/non-striped polymorphism in snakes (Reptilia: Ophidia). *Biological Reviews*, 69(4), pp.599-610.
- Xie, B., Song, X., 2022. The impaired unfolded protein-premelanosome protein and transient receptor potential channels-autophagy axes in apoptotic melanocytes in vitiligo. *Pigment Cell & Melanoma Research* 35, 6–17.. doi:10.1111/pcmr.13006
- Yin, W., Wang, Z.-J., Li, Q.-Y., Lian, J.-M., Zhou, Y., Lu, B.-Z., Jin, L.-J., Qiu, P.-X., Zhang, P., Zhu, W.-B., Wen, B., Huang, Y.-J., Lin, Z.-L., Qiu, B.-T., Su, X.-W., Yang, H.-M., Zhang, G.-J., Yan, G.-M., Zhou, Q., 2016. Evolutionary trajectories of snake genes and

- genomes revealed by comparative analyses of five-pacer viper. *Nature Communications* 7, 13107.. doi:10.1038/ncomms13107
- Zaher, H., Murphy, R.W., Arredondo, J.C., Graboski, R., Machado-Filho, P.R., Mahlow, K., Montingelli, G.G., Quadros, A.B., Orlov, N.L., Wilkinson, M., Zhang, Y.-P., Graziotin, F.G., 2019. Large-scale molecular phylogeny, morphology, divergence-time estimation, and the fossil record of advanced caenophidian snakes (Squamata: Serpentes). *PLOS ONE* 14, e0216148.. doi:10.1371/journal.pone.0216148
- Zanetti, G., Duregotti, E., Locatelli, C.A., Giampreti, A., Lonati, D., Rossetto, O., Pirazzini, M., 2018. Variability in venom composition of European viper subspecies limits the cross-effectiveness of antivenoms. *Scientific Reports* 8.. doi:10.1038/s41598-018-28135-0
- Zhang, H., Kranzler, H.R., Weiss, R.D., Luo, X., Brady, K.T., Anton, R.F., Farrer, L.A., Gelernter, J., 2009. Pro-Opiomelanocortin Gene Variation Related to Alcohol or Drug Dependence: Evidence and Replications Across Family- and Population-based Studies. *Biological Psychiatry* 66, 128–136.. doi:10.1016/j.biopsych.2008.12.021
- ZINENKO, O. I. Habitats of *Vipera berus nikolskii* in Ukraine. *Proceedings of the 13th Congress of the Societas Europaea Herpetologica*. pp, 2006. 209.

Supplementary Information

SI 1: DNA extraction

The DNeasy Blood & Tissue extraction kit from Qiagen (Hilden, Germany) was used to extract DNA from ~25 mg of shed skin, tissue, buccal swabs, or scale clippings from 236 *Vipera berus berus*, nine *Vipera berus bosniensis*, three *Vipera berus nikolskii*, three *Vipera ursinii ursinii*, three *Vipera ursinii macrops*, two *Vipera anatolica*, two *Vipera kaznakovi*, two *Vipera renardi*, one *Vipera graeca*, one *Vipera seoanei* in batches of 48, each with their own extraction blank. Each buccal swab was cut into strips before loading them into 1.5-ml Eppendorf tubes to maximize surface area for the extraction. The remaining samples were placed into 1.5-ml Eppendorf tubes untouched. 180 µL of buffer ATL and 20 µL of proteinase K were added in that order into each sample tube and vortexed before wrapping the tops with parafilm and incubating them at 56°C in a Thermoshaker set to 700 RPM for 24 hours. As the samples are incubating, a mixture of equal parts buffer AL and 96% ethanol was mixed to expedite the processing of multiple samples. Each tube was then vortexed briefly before adding 400 µL of the premixed buffer AL and ethanol solution and thoroughly vortexing again.

The entire contents of the sample tubes, avoiding large undigested tissue, were transferred to DNeasy Mini spin columns placed in 2 ml collection tubes and centrifuged at 8,000 rpm for 1 minute and placed into a new 2 ml collection tube. To wash the samples, 500 µL Buffer AW1 was added to each tube and centrifuged for one minute at 8,000 rpm. Afterward, the DNeasy Mini spin columns were transferred to the final two ml collection tube. 500 µL of Buffer AW2 was added for the final wash and then centrifuged for three minutes at 14,000 rpm to dry the DNeasy membrane. To ensure the silica membrane was dried, the flow-through was discarded after centrifuging and centrifuged a second time for one minute at 14,000 rpm.

The DNeasy Mini spin columns were transferred to a 1.5 ml microcentrifuge tube. To elute the DNA, 200 µL of Buffer AE was pipetted directly onto the DNeasy membrane and incubated for one minute at room temperature before centrifuging a final time for one minute at 8,000 rpm. The DNA yield was quantified using the Thermo Fisher Scientific Qubit 2.0 fluorometer (Indiana, USA) with the dsDNA HS (High Sensitivity) Assay kit.

SI 2: Library build protocol

To prepare the samples for shearing, they were diluted with EB buffer, if necessary, to achieve a concentration of 500 - 5000 ng DNA per 60 µL. After, the DNA samples were sheared using the Covaris ME220 focused ultrasonicator to reach a target fragmentation length of 400 bp. Some samples, however, were instead sheared using the Bioruptor Pico Sonication System from Diagenode Innovating Epigenetic Solutions with a target fragmentation length of 400 bp.

Blunt end repair

In batches of 95, 32 µL of each sample was loaded into a 96 well plate with one library blank of molecular-grade water in the 96th well. The master mix for repairing the

blunt ends and overhangs in the fragmented DNA was prepared over ice yielding 8 μL per reaction in addition to 8 extra reactions as the master mix is quite viscous when the reaction buffer is added. Per reaction, the master mix contained 0.4 μL T4 DNA polymerase at 0.03 U/ μL final concentration, 1 μL T4 polynucleotide kinase at 0.25 U/ μL , 0.4 μL dNTPs at 0.25 mM each, 4 μL T4 DNA ligase buffer (NEB) at 1x and 2.2 μL reaction buffer (PEG-4000 at 25% final concentration, BSA at 2mg/ML, NaCl at 400 mM and H₂O). Following the addition of the master mix, each sample was mixed by pipetting about 15 times, sealed with aluminum foil and spun down. The reactions incubated in the thermo-cycler for 30 minutes at 20 °C followed by 30 minutes at 65 °C. The reactions were then cooled to 4 °C.

Adapter ligation

To begin ligation of the adapters to the DNA, 2 μL of adapter solution (20 μM) was added to the samples and vortexed. 8 μL of master mix (T4 DNA ligase buffer at 1x, PEG-4000 at 6.25% final concentration and T4 DNA ligase (NEB- 400 U/ μL)) was also added and pipetted about 15 times to mix. The reactions were then incubated for 30 minutes at 20 °C, 10 minutes at 65 °C, and cooled to 4 °C.

Adapter fill-in

10 μL of master mix (2 μL Isothermal amp. Buffer (0.33x concentration), 0.8 μL dNTPs (0.33mM), 1.6 μL Bst 2.0 Warmstart pol. (0.21 U/ μL), 5.6 μL H₂O) was prepared over ice and added to each well. The reactions were mixed by pipetting up and down about 15 times and then incubated for 15 minutes at 65 °C followed by 15 minutes at 80 °C. The reactions were then cooled to 4 °C.

Library purification

100 μL of room temperature SPRI beads (SpeedBeads suspended in a solution composed of TE buffer, PEG-8000, NaCl, Tris_HCl, Tween20 and molecular grade water) were added to each library (60 μL ; Ratio Beads:Library: 1.67) and incubated for 5 minutes to allow the DNA to bind to the beads. The plate was then placed on a magnetic rack until pellets had formed. The supernatant was removed and the pellets were washed twice with 200 μL 80% Ethanol. To elute the DNA, the plate was removed from the magnetic rack and 33 μL of EB buffer was added to each sample, making sure to pipette up and down to fully re-suspend the beads. The samples incubated for 10 minutes at 37 °C and were placed back on the magnetic rack until pellets had formed. The supernatant was then transferred to a new plate.

qPCR

To determine the optimal number of PCR cycles during indexing, a qPCR was performed. The master mix was prepared on ice by combining 13.08 μL molecular grade water, 2 μL Taq AmpliTaq Gold buffer at 1x final concentration, 0.16 μL dNTPs at 0.2 mM, 0.4 IS7 at 0.2 μM , 0.4 IS8 at 0.2 μM , 2.0 μL MgCl₂ at 2.5 mM, 0.8 SYBR Green Mix and

0.16 μL AmpliTaq Gold polymerase at 0.4 U/ μL , in that order to yield 19 μL per reaction. In addition to the master mix, each well received 1 μL of diluted unamplified libraries, excluding the qPCR blank which received 1 μL of molecular grade water instead. The qPCR program started with 10 minutes at 95 $^{\circ}\text{C}$, then 40 cycles denature at 95 $^{\circ}\text{C}$ for 30 seconds, annealing for 60 seconds at 60 $^{\circ}\text{C}$ and 45 seconds at 72 $^{\circ}\text{C}$. To finish the qPCR there was also a melt curve.

Indexing PCR and sequencing

Each sample was indexed with a unique combination of forward and reverse index primers. Combinations were mapped out previously to avoid duplicate combinations. The master mix was prepared on ice and all reagents were vortexed briefly before use, excluding the enzymes. 86 μL index PCR master mix (62.2 μL H₂O, 10 μL AmpliTaq Gold buffer 1x, 0.8 μL dNTPs at 0.2 mM final concentration, 2 μL BSA at 0.4 mg/ml, 10 μL MgCl₂ at 2.5 mM, 1 μL AmpliTaq Gold polymerase at 0.05 U/ μL) was pipetted into each plate well in addition to 10 μL undiluted library template, 2 μL forward index primer and 2 μL reverse index primer, resulting in 100 μL total volume for each sample. Based on the results of the qPCR, samples were grouped together that required the same number of cycles \pm 1 or 2 cycles. The PCR program started with 10 minutes at 95 $^{\circ}\text{C}$, followed by 30 seconds at 95 $^{\circ}\text{C}$ to denature, 60 seconds at 60 $^{\circ}\text{C}$ to anneal and 45 seconds at 72 $^{\circ}\text{C}$ repeated for the determined number of cycles for each group. To end the PCR program, all groups had a final extension for 5 minutes at 72 $^{\circ}\text{C}$ followed by a drop in temperature to 4 $^{\circ}\text{C}$ until the samples were retrieved from the thermocycler.

Purifying and pooling indexed PCRs with SPRI beads







The amplified libraries were purified using 100 μL of SPRI beads in each plate well and pipetted up and down to fully incorporate. The rack was then placed onto the magnetic rack to incubate at room temperature for 5 minutes until the bead pellet had formed. The supernatant was removed and the pellets were washed twice with 200 μL 80% ethanol before removing the plate from the magnetic rack. 33 μL of EB buffer was added to each library of the plate to elute the beads and DNA. The plate was then sealed, vortexed and incubated at 37 $^{\circ}\text{C}$ for 10 minutes. After incubation, the plates were placed back onto the magnetic rack for the final time until pellets had formed. Once the pellets had formed, the supernatant was transferred into a new plate as the purified, amplified and indexed libraries.

To design the pools, the libraries were run on the Agilent TapeStation 4200 to estimate molarity and verify fragment size. The molarity results from the TapeStation 4200 were used to calculate the libraries that needed to be diluted. The same amount of each library was added (concentration and volume) to the pools to reach 5nM concentration or higher.

Supplementary figures

SF 1: Isohelix buccal swab instructions

Instructions for Isohelix SK-1S, SK-3S, SK-4S and SK-5S Buccal Swabs

<p>1</p> 	<p>Pull open the package from one end.</p>
<p>2</p> 	<p>Remove the swab from the tube, taking care not to touch the white swab head with your fingers</p>
<p>3</p> 	<p>Insert the swab into your mouth and rub firmly against the inside of your cheek or underneath lower and upper lip. For standard DNA collection rub for 1 minute and in all cases rub for a minimum of 20 seconds. Important – use reasonable, firm and solid pressure</p>
<p>4</p> 	<p>Place the swab back into the tube. Do not touch the swab head with your fingers</p>
<p>5</p> 	<p>Place your thumbnail in the small groove set in the handle, then snap the handle in two by bending to one side. Let the swab head fall into the tube.</p>
<p>6</p> 	<p>Seal the tube securely with the cap provided.</p>