**Title**

# Shed skin as a source of DNA for genotyping-by-sequencing (GBS) in reptiles

**Running title**

Shed skin GBS

**Authors**

Thomas D Brekke[1], Liam Shier[1], Matthew J Hegarty[2], John F Mulley[1]*

**Affiliations**

1. School of Natural Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, United Kingdom
2. Institute of Biological, Environmental & Rural Sciences, Aberystwyth University, Penglais, Aberystwyth, Ceredigion, SY23 3DA, United Kingdom

* To whom correspondence should be addressed: j.mulley@bangor.ac.uk

**Keywords**

Snake, reptile, GBS, skin, SNP

## Abstract

Association and genetic mapping studies aimed at linking genotype to phenotype are powerful tools that require large numbers of samples, complicating their use in long-lived species with low fecundity. Shed skins of snakes and other reptiles contain DNA; are a safe and ethical way of non-invasively sampling large numbers of individuals; and provide a simple mechanism by which to involve the public in scientific research. Here we test whether the DNA in dried shed skins mailed to us from citizen scientists is suitable for reduced representation sequencing approaches, specifically genotyping-by-sequencing (GBS). We find that shed skin samples provide DNA of sufficient quality and quantity for GBS, although libraries from shed skin resulted in fewer sequenced reads than libraries from snap-frozen muscle, and contained slightly fewer variants (70,685 SNPs versus 97,724). This issue is a direct result of lower read counts of the shed skin samples, and can be rectified quite simply with deeper sequencing. Skin-derived libraries also have a very slight (but significantly different) profile of transitions and transversions, suggesting that DNA damage occurs but is minimal. We conclude that shed skin-derived DNA is a good source of genomic DNA for a variety of genetic studies, and use it to identify sex-linked scaffolds in the corn snake genome.

## Introduction

Animal pigmentation fulfils a variety of functions, including intra- and inter-specific recognition and communication, thermoregulation, crypsis, masquerade, and mimicry (Kronforst et al., 2012; Olsson, Stuart-Fox, & Ballen, 2013; Rosenblum, Hoekstra, & Nachman, 2004). The variation in pigmentation patterning on different parts of the body within or between species represents an excellent opportunity with which to identify the genetic and developmental basis of morphological variation, and to relate this to ecological and evolutionary pressures. Snakes represent excellent models with which to study vertebrate pigmentation patterning, as different species, and even different populations of the same species, demonstrate a diverse range of patterns (spots, longitudinal and lateral stripes, and complex patterns not found on other species) but share a common body plan. The simplicity of this body plan facilitates mathematical modelling of pattern formation processes (Murray & Myerscough, 1991). Snake patterns play a vital role in thermoregulation, camouflage (both crypsis and masquerade), and defence, through warning

and startle colouration, mimicry, and differential head/body colour and pattern, which confuses the head-body boundary. In addition to the wide diversity of pigmentation patterns seen in nature, there are a huge number of pigment and pattern variants available in the commercial pet trade (so-called "morphs"). To their owners these are simply attractive pets, but to a biologist they are a useful group of mutants with defects in pigment biosynthesis, and/or neural crest formation, differentiation or migration, that will allow us to identify the genetic and developmental basis of intra-specific variation.

Linking phenotypes to their underlying genetic basis requires genomic resources, and although there are now several snake genomes (Castoe et al., 2013; J.-T. Li et al., 2018; Perry et al., 2018; Shibata et al., 2018; Ullate-Agote, Milinkovitch, & Tzika, 2014; Vonk et al., 2013; Yin et al., 2016), many of these are poor quality, and all represent a single individual. Reduced representation sequencing approaches such as restriction site-associated DNA sequencing (RADseq) (Baird et al., 2008; Davey & Blaxter, 2010; Hohenlohe et al., 2010; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) and genotyping-by-sequencing (GBS) (Elshire et al., 2011; Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013) therefore offer the best opportunity for genome-scale analysis from large number of individuals, and particularly to associate genetic markers with phenotypes of interest. High quality DNA can be obtained from snakes and other reptiles in a number of ways, including blood (where nucleated red blood cells facilitate lower sample volumes than would be required in mammals) and tissue (e.g. scale, toe, or tail clips (Beebee, 2008; Maigret, 2018)). However, invasive sample collection can be technically and logistically challenging, and carries with it ethical and animal welfare concerns. Whilst non-invasive approaches are therefore increasingly favoured, they are not without their own difficulties. Chemical preservation of museum specimens can degrade DNA, and procurement of adequate samples from roadkill specimens (whilst often high quality if found soon after death) is sporadic and unpredictable. Fecal samples inevitably contain large amounts of microorganisms, and DNA often degrades quickly unless samples are rapidly frozen (Jones, Cable, & Bruford, 2008), and cloacal and buccal swabbing (Beebee, 2008; Miller, 2006; Pidancier, Miquel, & Miaud, 2003) is dependent on locating and restraining animals, with research on venomous animals carrying particular risks. Shed skins therefore have great potential for non-invasive sampling of reptile DNA, without the need to actually locate and handle/restrain the animal itself, and without associated ethical and animal welfare issues. Such risk-

free approaches lend themselves especially to citizen science projects, where members of the public can collect and ship samples from a far larger area and at far lower cost. DNA derived from shed skins of lizards and snakes has long been known to be of sufficient quality and quantity for PCR-based genotyping of a small number of genetic markers (Bricker, Bushar, Reinert, & Gelbert, 1996; Fetzner Jr, 1999; Horreo, Peláez, & Fitze, 2015; Tawichasri et al., 2017; Villarreal, Bricker, Reinert, Gelbert, & Bushar, 1996), and so here we assess the utility of shed skin for larger-scale single nucleotide polymorphism (SNP) genotyping using genotyping-by-sequencing (GBS).

## Materials and Methods

### *DNA extraction*

Shed skins were collected from 61 corn snakes (*Pantherophis guttatus*) from our in-house colony, and from commercial breeders and hobbyist keepers. Skins were collected as soon as possible after shedding, and stored at -20°C. Those collected for us by others were placed into individual paper envelopes for shipping and were stored at -20°C upon arrival in Bangor. DNA was extracted from around 50mg samples of ventral scale skin using the DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's protocol, with the exception of a longer (24 hour) proteinase treatment at 56°C. Small DNA fragments were removed by spin-column chromatography with Chroma-Spin-1000+TE columns (Clontech) following the manufacturer's protocol, and samples were quantified using the Qubit dsDNA BR assay kit and Qubit Fluorometer. In some cases, it was necessary to perform multiple extractions from a single sample and pool them using ethanol precipitation to obtain the desired 100ng of DNA. We also prepared DNA from 50mg samples of snap-frozen muscle from 18 individuals using the same procedure.

### *Genotyping-by-sequencing*

Libraries were prepared as described by Elshire et al. (Elshire et al., 2011), at the Institute of Biological, Environmental and Rural Sciences (IBERS) at Aberystwyth University. Restriction digestion was carried out using the type II restriction endonuclease PstI, and the resulting fragments were tagged with 81 unique barcodes of varying lengths (two of the 79 individuals were given two barcodes and run as two

separate samples totalling 81 barcodes). GBS libraries were pooled to an equimiolar concentration and single-end sequenced on one lane of Illumina HiSeq 2500.

*Bioinformatics*

We used the stacks (v1.44) pipeline (J. Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; J. M. Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011) to do a genome-guided stacks assembly and call SNPs. Our pipeline started with the program 'process_radtags' and took as arguments the single fastq file (-f), the list of barcodes (-b), the restriction enzyme (-e pstI), as well as the flags to clean the data (-c), discard reads with low quality (-q), rescue the barcodes (-r), specify the quality encoding (-E phred33), and specify how the barcodes were situated in the reads (--inline_null). Once process_radtags had finished de-multiplexing and cleaning the data, we counted the remaining high-quality reads for each barcode and then merged the two duplicate individuals. We aligned all reads to the corn snake genome (GCA_001185365.1 (Ullate-Agote et al., 2014)) with BWA (H. Li & Durbin, 2009) and counted the depth of coverage with 'samtools depth' (H. Li et al., 2009). We processed the genome alignments with 'pstacks' using a minimum stack depth of 3 (-m 3), and then built the catalog with 'cstacks' using all individuals and allowing for 1 mismatch against the reference (-n 1 ), and then ran 'sstacks' with all default settings. Finally, we treated all individuals as belonging to a single group and ran 'populations' to call variants. We included flags to keep SNPs that are present in a single population (-p 1), required a minimum of 5 reads to call a stack at a locus (-m 5), kept only SNPs (--remove-indels), and export the variant calls in vcf format (--vcf). We then filtered the variant file using Vcftools v0.1.15 (Danecek et al. 2011) for sites with more than 60% completeness (--max_missing .6), calculated the percent missing genotypes for each individual (--missing-indv), and the transition-transversion profiles (--TsTv-summary) for both skin and muscle libraries separately, and then calculated the genome-wide Weir and Cockerham mean Fst between the skin- and muscle-derived libraries (--weir-fst-pop).

*Identification of sex chromosome-specific markers*

We used the coverage from each individual to identify sex-linked genomic contigs as described previously (Brekke et al., 2019; Brekke, Steele, & Mulley, 2018) by first calculating each individuals' sequencing effort as the sum of aligned reads for that individual. We standardised the contig-level counts by dividing by the sequencing effort of each individual and multiplying by 1,000,000. We compared the mean of the coverage of females and the mean coverage of males for each contig. W contigs should be present in females but not males and fulfil the inequality:

$$Coverage_{Male} < 0.1 * Coverage_{Female} - 0.2$$

Unknown contigs are the not W-linked and have overall standardized coverage of less than 5:

$$Coverage_{Male} + Coverage_{Female} < 5$$

Z-linked contigs are not unknown and have twice the coverage in males as females and fulfil the inequality:

$$Coverage_{Male} < 1.5 * Coverage_{Female} + 0.2$$

All remaining contigs are annotated as autosomal. These specific cut-offs were chosen based on the natural breakpoints in the plot.

**Results**

DNA was successfully extracted from all skin and muscle samples, with concentrations varying between 0.37-12ng/µl (mean 6.15ng/µl). In all cases, only a small proportion of a shed skin (typically less than 200mg) was needed to obtain sufficient DNA for GBS. On average there were 2,818,124 ± 1,154,007 reads sequenced per individual. GBS libraries extracted from muscle tissue had more sequenced reads than libraries from skin tissue (Figure 1, muscle mean: 3,798,510 reads, skin mean: 2,538,013 reads, Welch two sample T-test, t=4.4064, df=26.183, *P*=0.0001589). We identified 237,466 total raw SNPs. After filtering for completeness we found 101,618 SNPs: 97,724 SNPs at >60% completeness in the muscle samples and 70,685 with >60% completeness in the skin samples, of which 66,791 were found in both muscle and skin. Even after filtering for overall missing data, the genotyping rate was highly variable across samples with an average of 16.7% ± 12.7% missing in any given sample and muscle-derived libraries had fewer missing genotypes than skin-derived libraries (Figure 2, muscle mean: 11.3% missing, skin mean: 18.2% missing, Welch two sample T-test, t = -3.267, df = 74.723, *P* = 0.001643). Furthermore, there is a strong relationship

between the amount of sequencing and the genotyping rate, especially at read counts lower than 1,000,000 (Figure 3).

Low read counts and much missing data may suggest that the DNA has been damaged prior to library construction. To test for DNA damage we analysed the distribution of transitions and transversions and found significant differences between the skin- and muscle-derived libraries in terms of the number of transitions and transversions (Figure 4, Chi square test: $X^2 = 15.843$, df = 5, $P = 0.007306$). While the differences are significant, the effect sizes are slight. The transition to transversion ratio for muscle-derived libraries is 2.652 and for skin-derived libraries it is 2.714 and within each class the two library types differ by only tenths of a percentage: AC in skin is 6.92% and in muscle is 7.07%, AT is 6.57% in skin versus 6.41% in muscle, CG is 6.87% versus 6.92%, GT is 6.86% versus 6.66%, AG is 36.37% versus 36.76%, and CT is 36.24% versus 36.31%. In addition, the Fst between the skin and muscle samples is quite low (Fst = 0.0686) suggesting that there is little genome-wide differentiation between skin and muscle samples.

We used coverage to identify the sex-linked scaffolds in the genome (Figure 5) and were able to reliably annotate approximately 30% of the genome. 58,935 scaffolds (349,080,366 bases) are autosomal, 4178 scaffolds (19,511,706 bases) are Z-linked, and 1275 scaffolds (2,357,273 bases) are W-linked, the remaining 819,528 scaffolds (1,033,270,996 bases) had too little coverage to discriminate chromosomal linkage (Supplemental file 1).

## Discussion

DNA extracted from shed skins is suitable for use with reduced-representation sequencing approaches such as GBS with some considerations. We compared GBS libraries built from skin-extracted DNA with libraries from muscle-extracted DNA. The muscle samples were immediately snap-frozen and are thus a source of high-quality DNA while the skins were collected by pet snake owners across the country, dried, and shipped at ambient temperature through the post and thus subject to a variety of mechanisms of DNA degradation. Accordingly, we found significant differences between the library preparations of muscle and skin-derived DNA that likely stem from DNA damage in the original samples. These differences include the number of reads sequenced (Figure 1) and the SNPs identified (Figure 4) by each library type.

However, dealing with the differences between the two library types is surmountable with some forethought toward experimental design.

The number of SNPs identified in a GBS experiment depends strongly on the sequencing depth (Figure 2). Thus sequencing skin samples more deeply than samples from fresh tissue will help assure that a sufficient number of real SNPs can be identified. Estimating the necessary coverage in a GBS experiment is difficult as it depends on the frequency of cut-sites in the genome (often unknown in non-model species), the heterozygosity present in the sample population (often unknown until at least after the first round of sequencing), the specific cut-offs used to filter variant sites, and the type of experiment (population genetics may require different number of SNPs than parentage analyses etc). For this experiment the SNP discovery curve plateaued at around 3,000,000 reads per individual (Figure 3), and many of the skin-derived libraries had far fewer genotyped SNPs due to low read coverage (Figures 2 and 3). As such, we suggest that researchers plan on sequencing skin and other possibly-damaged samples more deeply in order to identify a robust set of variants.

Some problems, especially those relating to DNA damage, cannot be dealt with simply by sequencing more deeply. For instance, if Cytosine deamination into Uracil is a common issue in the sample (as is often the case in ancient DNA studies (Hofreiter, Serre, Poinar, Kuch, & Pääbo, 2001), more sequencing will not remove those errors and a more sophisticated approach is needed. The signal of errors in the skin libraries is very slight suggesting that identifying and dealing with these errors would be difficult. Fortunately such labour-intensive error cleaning will likely provide little benefit for three reasons. First, the vast majority of the SNPs identified in the skin samples are also found in the muscle samples (Figure 4a) and this is true despite the excess of skin-derived libraries (61 vs 18 muscle-derived) in which to discover SNPs. Even if the 3,894 skin-specific SNPs are not biologically real, they are so few (only 3.8% of all SNPs) that identifying them in the absence of other high quality libraries will be exceedingly difficult and, more importantly, removing them will have little effect on the overarching results. Secondly, while the skin and muscle may have slightly different SNP profiles (Figure 4b), the effect is so slight that many draconian filtering options (such as removing all C/T sites or ignoring all transversions) are not merited. Finally if there was a systematic damage pressure it should alter the allele frequencies in the damaged samples, but the

Fst between the skin and muscle samples is very low (genome-wide mean Fst is 0.0686. While this metric can only be calculated for the 66,791 shared SNPs, it shows that there are not striking allele frequency differences between skin- and muscle-derived libraries which is evidence that the degradation is not a major concern. Such a low Fst result is additionally convincing given that the individuals from whom muscle was used are all close relatives in our colony while the snakes whose skins were used originate from breeders across the UK. This confounding population structure should artificially increase our estimate of Fst, and the Fst due solely to DNA damage in the skin is likely much smaller. In sum, while DNA damage is apparent in our skin samples, the effects are slight and not likely to impact any biological results.

Shed skins are a safe and ethical way to collect DNA, especially from wild or venomous species, and thus facilitate the engagement of the public in sample collection. Public sourcing of sample material is crucially important for genetic studies in long-lived and slow-breeding animals like snakes as the time to collect sufficient samples for association studies or breed sufficient animals for a genetic mapping experiment can be prohibitive. Leveraging the amateur herpetological community and sourcing skins from the public has the dual benefit of engaging citizen scientists and the possibility of rapidly collecting extremely large sample sizes. Pet snakes in general, and corn snakes in particular have a variety of colour and pattern morphs which may prove incredibly powerful for understanding the genetic basis for colouration. We have shown that a simple crowd-sourced collection technique (mailing shed snake skins), can provide samples containing DNA of sufficient quality for reduced representation sequencing. This finding opens up the possibility of doing association studies on patterning and colouration from a wealth of samples in a long-lived and low-fecundity species.

We also used the relative coverage of all scaffolds in males compared to females to identify and annotate the sex-linked scaffolds in corn snakes. This approach does not order the scaffolds into complete chromosomes but it does discriminate between Z- and W- linked scaffolds which represents a marked improvement in the corn snake genome (Figure 5 and Supplemental File 1). This information facilitate the design of sex-specific PCR primers which can be used on DNA extracted from shed skins to non-invasively sex-type snakes.

239

240 **Acknowledgements**

241 We wish to thank Adam Clarke and numerous anonymous donors for both financial support and
242 donation of shed skins. We also wish to thank the School of Natural Science technical team, and especially
243 Rhys Morgan, for assistance with animal care, and Adam Hargreaves for useful discussions.

244

245 **References**

246 Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008).
247 Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One*, *3*(10), e3376.
248 Beebee, T. J. C. (2008). Buccal swabbing as a source of DNA from squamate reptiles. *Conservation Genetics*,
249 *9*(4), 1087–1088.
250 Brekke, T. D., Steele, K. A., & Mulley, J. F. (2018). Inbred or Outbred? Genetic diversity in laboratory rodent
251 colonies. *G3: Genes, Genomes, Genetics*, *8*(2), 679–686.
252 Brekke, T. D., Supriya, S., Denver, M. G., Thom, A., Steele, K. A., & Mulley, J. F. (2019). A high-density
253 genetic map and molecular sex-typing assay for gerbils. *Mammalian Genome*, *30*(3–4), 63–70.
254 Bricker, J., Bushar, L. M., Reinert, H. K., & Gelbert, L. (1996). Purification of high quality DNA from shed
255 skins. *Herpetological Review*, *27*(3), 133.
256 Castoe, T. A., de Koning, A. P. J., Hall, K. T., Card, D. C., Schield, D. R., Fujita, M. K., … Pollock, D. D. (2013).
257 The Burmese python genome reveals the molecular basis for extreme adaptation in snakes.
258 *Proceedings of the National Academy of Sciences of the United States of America*, *110*(51), 20645–
259 20650.
260 Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set
261 for population genomics. *Molecular Ecology*, *22*(11), 3124–3140.
262 Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: building and
263 genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*(3), 171–182.
264 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & McVean, G. (2011). The
265 variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.

Davey, J. W., & Blaxter, M. L. (2010). RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, *9*(5–6), 416–423.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, *6*(5), e19379.

Fetzner Jr, J. W. (1999). Extracting high-quality DNA from shed reptile skins: a simplified method. *Biotechniques*, *26*(6), 1052–1054.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., & Pääbo, S. (2001). ancient DNA. *Nature Reviews Genetics*, *2*(5), 353.

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

Horreo, J. L., Peláez, M. L., & Fitze, P. S. (2015). Skin sheds as a useful DNA source for lizard conservation. *Phyllomedusa: Journal of Herpetology*, *14*(1), 73–77.

Jones, R., Cable, J., & Bruford, M. W. (2008). An evaluation of non-invasive sampling for genetic analysis in northern European reptiles. *The Herpetological Journal*, *18*(1), 32–39.

Kronforst, M. R., Barsh, G. S., Kopp, A., Mallet, J., Monteiro, A., Mullen, S. P., … Hoekstra, H. E. (2012). Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. *Pigment Cell & Melanoma Research*, *25*(4), 411–433.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, J.-T., Gao, Y.-D., Xie, L., Deng, C., Shi, P., Guan, M.-L., … Ding, L. (2018). Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. *Proceedings of the National Academy of Sciences*, *115*(33), 8406–8411.

293 Maigret, T. A. (2018). Snake scale clips as a source of high quality DNA suitable for RAD sequencing.
294 *Conservation Genetics Resources*, 1–3.
295 Miller, H. C. (2006). Cloacal and buccal swabs are a reliable source of DNA for microsatellite genotyping of
296 reptiles. *Conservation Genetics*, *7*(6), 1001–1003.
297 Murray, J. D., & Myerscough, M. R. (1991). Pigmentation pattern formation on snakes. *Journal of*
298 *Theoretical Biology*, *149*(3), 339–360.
299 Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013).
300 Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, *22*(11), 2841–
301 2847.
302 Olsson, M., Stuart-Fox, D., & Ballen, C. (2013). Genetics and evolution of colour patterns in reptiles.
303 *Seminars in Cell & Developmental Biology*, *24*(6–7), 529–541.
304 Perry, B. W., Card, D. C., McGlothlin, J. W., Pasquesi, G. I. M., Adams, R. H., Schield, D. R., … Hoffmann, F. G.
305 (2018). Molecular adaptations for sensing and securing prey and insight into amniote genome
306 diversity from the garter snake genome. *Genome Biology and Evolution*, *10*(8), 2110–2129.
307 Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an
308 inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS*
309 *One*, *7*(5), e37135.
310 Pidancier, N., Miquel, C., & Miaud, C. (2003). Buccal swabs as a non-destructive tissue sampling method for
311 DNA analysis in amphibians. *Herpetological Journal*, *13*(4), 175–178.
312 Rosenblum, E. B., Hoekstra, H. E., & Nachman, M. W. (2004). Adaptive reptile color variation and the
313 evolution of the Mc1r gene. *Evolution; International Journal of Organic Evolution*, *58*(8), 1794–1808.
314 Shibata, H., Chijiwa, T., Oda-Ueda, N., Nakamura, H., Yamaguchi, K., Hattori, S., … Isomoto, A. (2018). The
315 habu genome reveals accelerated evolution of venom protein genes. *Scientific Reports*, *8*(1), 11300.
316 Tawichasri, P., Laopichienpong, N., Chanhome, L., Phatcharakullawarawat, R., Singchat, W., Koomgun, T., …
317 Muangmai, N. (2017). Using blood and non-invasive shed skin samples to identify sex of caenophidian
318 snakes based on multiplex PCR assay. *Zoologischer Anzeiger*, *271*, 6–14.
319 Ullate-Agote, A., Milinkovitch, M. C., & Tzika, A. C. (2014). The genome sequence of the corn snake

(*Pantherophis guttatus*), a valuable resource for EvoDevo studies in squamates. *The International Journal of Developmental Biology*, *58*(10-11–12), 881–888. https://doi.org/10.1387/ijdb.150060at

Villarreal, X., Bricker, J., Reinert, H. K., Gelbert, L., & Bushar, L. M. (1996). Isolation and characterization of microsatellite loci for use in population genetic analysis in the timber rattlesnake, *Crotalus horridus*. *Journal of Heredity*, *87*(2), 152–155.

Vonk, F. J., Casewell, N. R., Henkel, C. V, Heimberg, A. M., Jansen, H. J., McCleary, R. J., … Richardson, M. K. (2013). The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(51), 20651–20656. https://doi.org/10.1073/pnas.1314702110 [doi]

Yin, W., Wang, Z., Li, Q., Lian, J., Zhou, Y., Lu, B., … Zhou, Q. (2016). Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. *Nature Communications*, *7*, 13107. https://doi.org/10.1038/ncomms13107

## Author Contributions

JFM devised the study. LS and MJH performed experiments, and TDB and JFM analysed the data. JFM and TDB wrote the paper, and all authors read and approved the final manuscript.

## Data Accessibility

Genotyping-by-sequencing data have been deposited in the European Nucleotide Archive, under accession numbers xxxx and Bioproject xxxx.

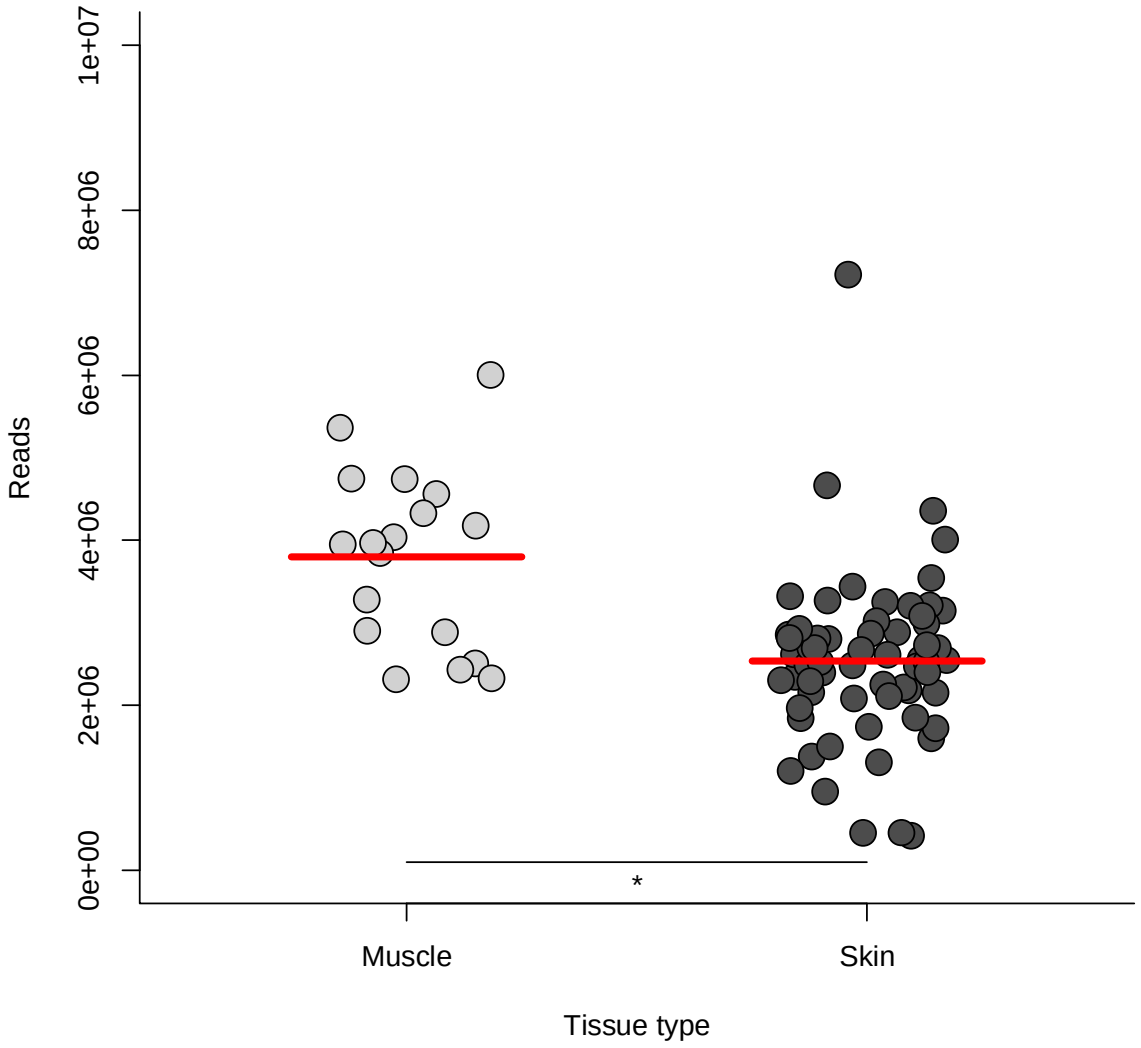



**Figure 1:** Read counts and genotyping completeness depend on tissue origin of the GBS library. Each sample is plotted according to its tissue of origin and horizontal lines denote the mean. Muscle-derived libraries have more reads than skin-derived libraries despite being pooled in equimolar ratios during pooling (Welch two sample T-test, t=4.4064, df=26.183, *P*=0.0001589). This pattern is likely caused by lower quality DNA from dried shed skins than snap-frozen muscle tissue.

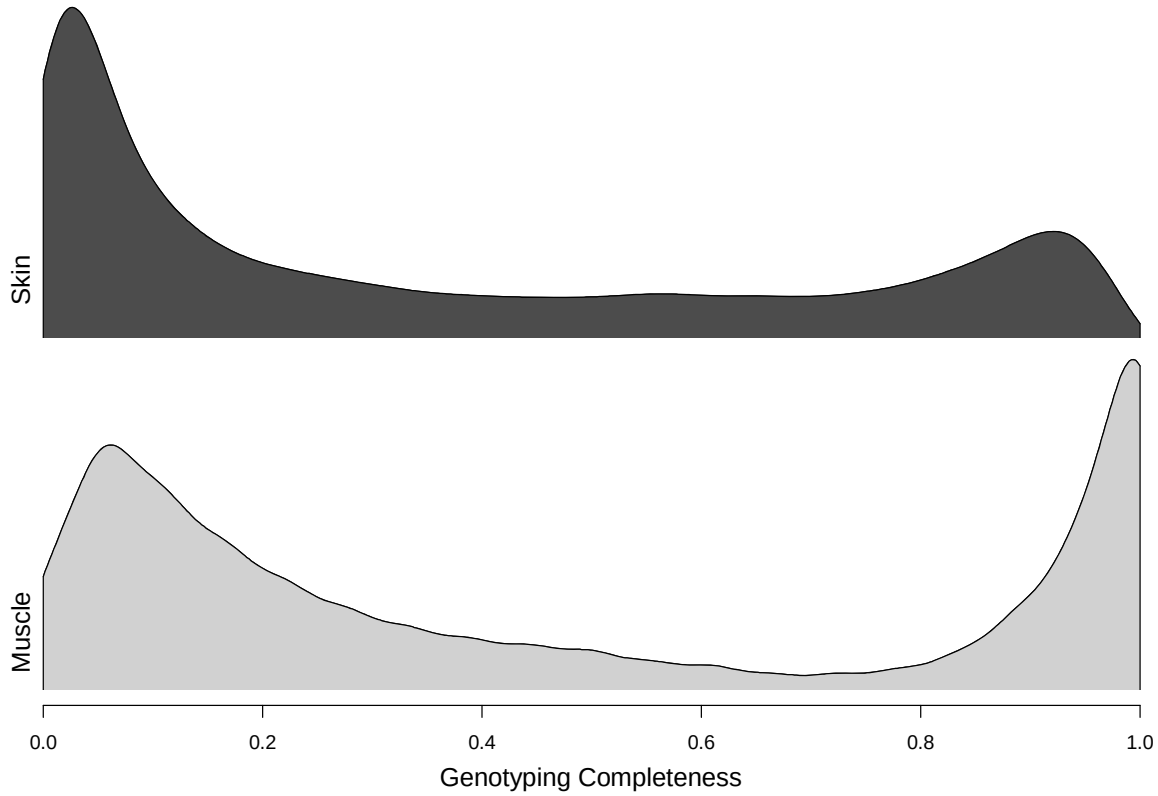
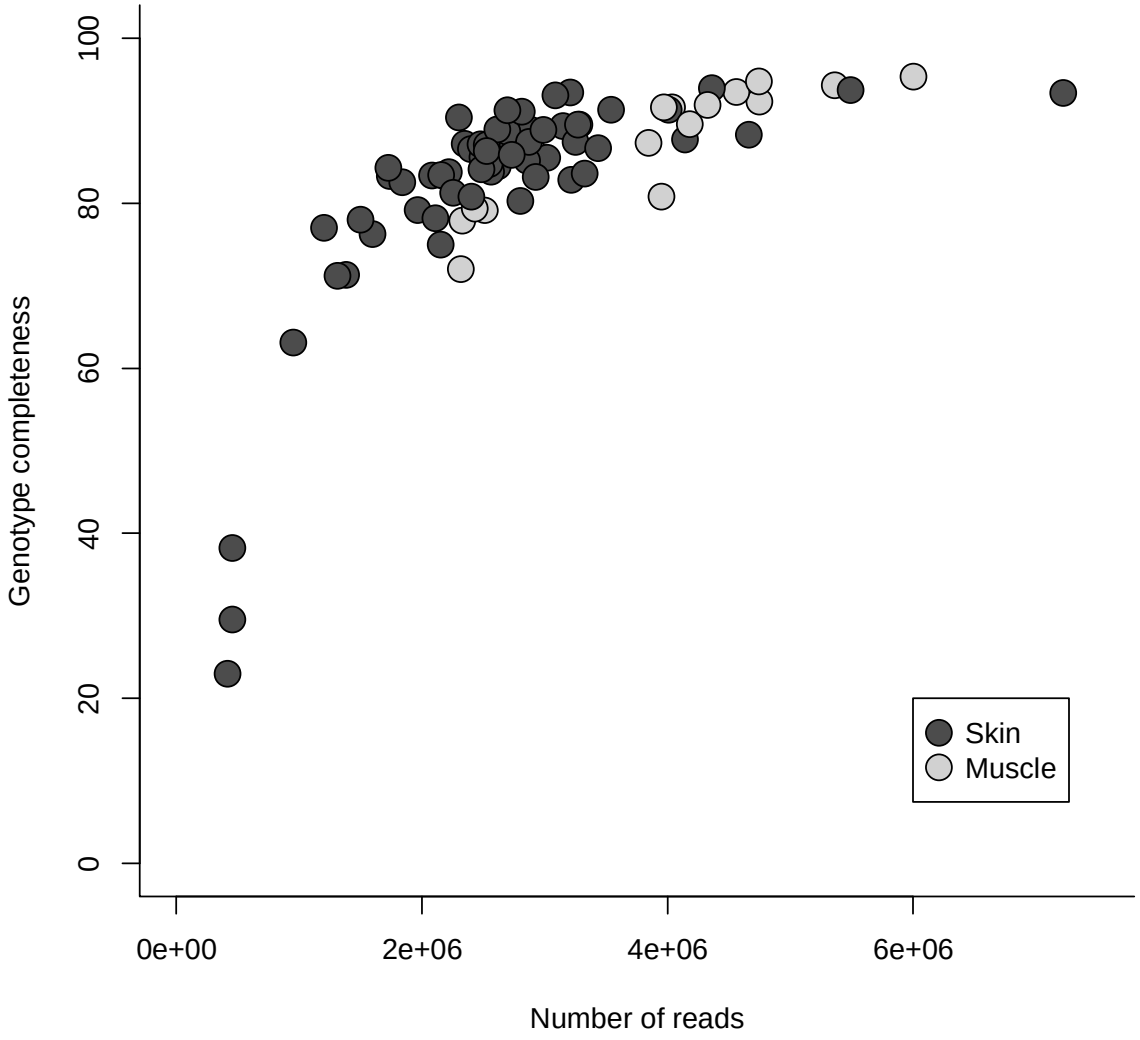



**Figure 2:** Fewer sites are genotyped in skin-derived libraries. These are density distributions showing the proportion of libraries for which each of the 237,466 raw SNPs are genotyped. At completeness of 0 the SNP is genotyped in no library, while SNPs at completeness of 100 are genotyped in all libraries. In both skin- and muscle-derived libraries there are many SNPs that are only genotyped in a few samples (peaks near 0), but muscle-derived libraries have far more SNPs that are genotyped in many samples (peaks near 1) (T-test, t = 119.47, df = 566235, *P* = 2.2e-16). All further analyses are done on the 101,618 SNPs at completeness of 60% and above.




366

367 **Figure 3:** There is a strong relationship between the number of reads sequenced per sample and the

368 number of missing genotypes per the individual such that with deeper sequencing more variants can be

369 called. There is a sharp cutoff at around 1,000,000 reads under which the missing genotype rate rises

370 sharply, and a consistently high genotyping rate does not occur until above 2,000,000 reads per sample.

371 This figure shows the subset of SNPs at the >60% completeness cutoff for each library type.

372

373

374

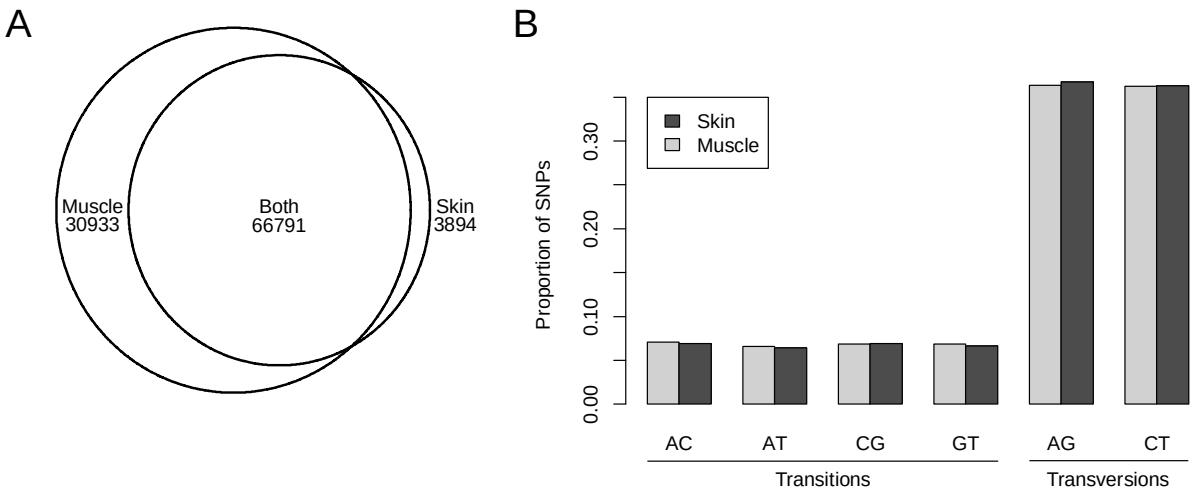
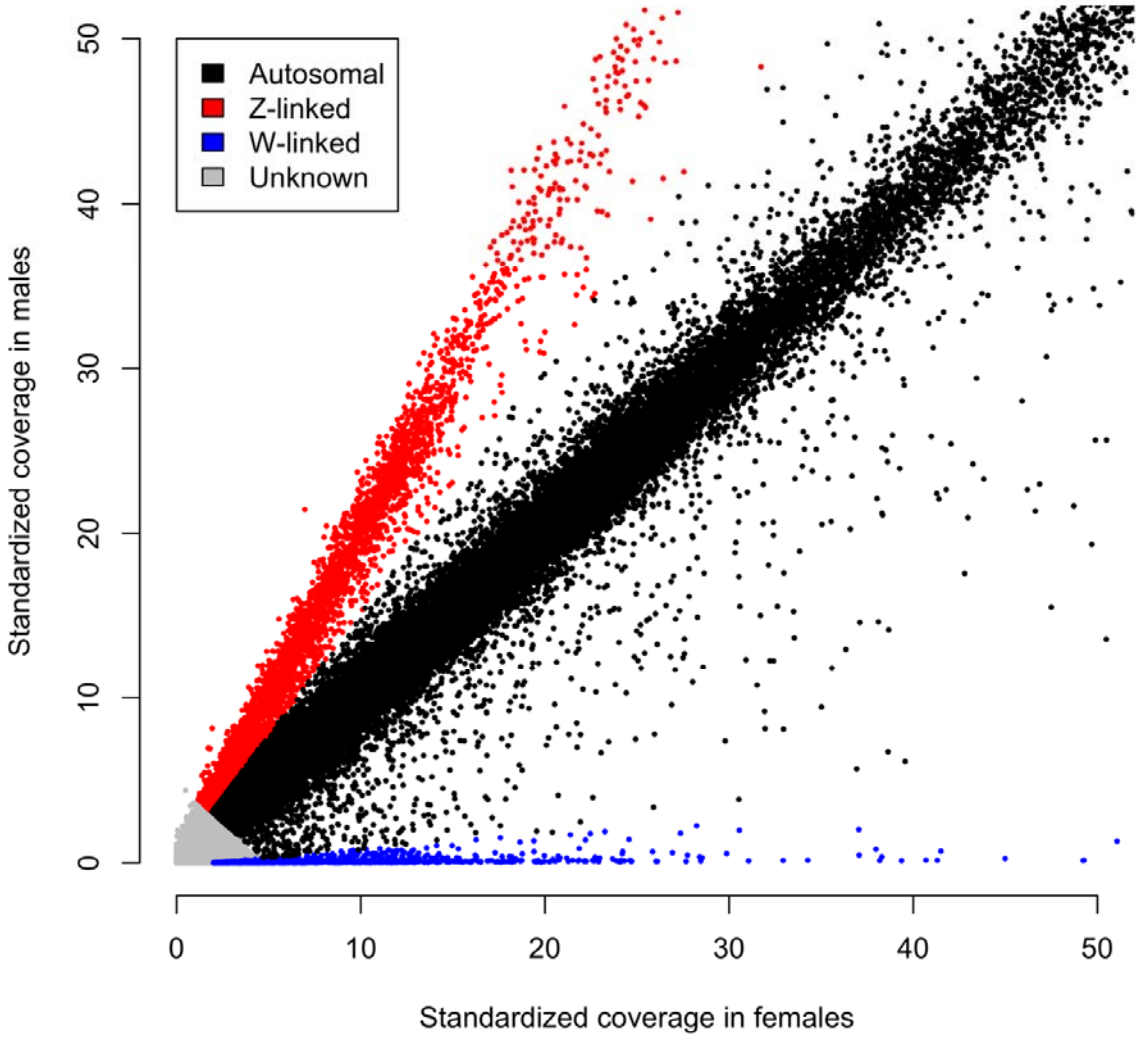



**Figure 4:** There are slight differences in SNP profiles between skin and muscle libraries. After filtering each set of libraries independently for 60% completeness, most SNPs were found in both muscle and skin (A). There are significant differences in the proportions of all SNP types (transitions: AC, AT, CG, and GT, and transversions: AG and CT) between skin and muscle (B, Chi square test: $X^2 = 15.843$, df = 5, $P = 0.007306$) but the effect size is vanishingly small. This suggests that DNA damage occurs in the dried skin, but that identifying and accounting for specific errors will be difficult and provide little benefit.




392

393 **Figure 5:** Relative coverage in males and females discriminates sex-linked from autosomal scaffolds. Points

394 on the 1:1 line have equal sequencing in males and femals and are thus autosomal. Z-linked scaffolds (red

395 points) and W-linked scaffolds (blue points) are also readily apparent. Grey points are scaffolds with too

396 little coverage to reliably discriminate relative coverage in males and females.

397

398

399