Methods and resource article

Word count 3990

Title

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

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

ORCID details

Thomas D Brekke 0000-0003-4479-1847

Matthew J Hegarty 0000-0001-6547-3800

John F Mulley 0000-0002-1537-7316

Abstract

Over a fifth of reptile species are classified as 'Threatened' and conservation efforts, especially those aimed at recovery of isolated or fragmented populations, will require genetic and genomic data and resources. 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 skin is suitable for reduced representation sequencing approaches, specifically genotyping-by-sequencing (GBS). Shed skin-derived libraries resulted in fewer sequenced reads than those from snap-frozen muscle samples, and contained slightly fewer variants (70,685 SNPs versus 97,724), but this issue can easily be rectified with deeper sequencing of shed skin-derived libraries. Skin-derived libraries also have a very slight (but significantly different) profile of transitions and transversions, most likely as a result of DNA damage, but the impact of this is minimal given the large number of single nucleotide polymorphisms (SNPs) involved. SNP density tends to scale with chromosome length, and microchromosomes have a significantly higher SNP density than macrochromosomes, most likely because of their higher GC content. Overall, shed skin provides DNA of sufficient quality and quantity for the identification of large number of SNPs, but requires greater sequencing depth, and consideration of the GC richness of microchromosomes when selecting restriction enzymes.

Keywords

SNPs; microchromosomes; corn snake; genetic diversity; population genetics

Introduction

The first extinction-risk assessment of reptiles was published in the spring of 2022, and showed that just over a fifth (21.1%) of reptile species are threatened, making them the second most vulnerable group after amphibians (Cox et al., 2022; 'The IUCN Red List of Threatened Species'). Conserving these species will require interventions at a variety of levels, including not only habitat preservation and restoration, but also the development of data and resources for the estimation of genetic factors associated with extinction risk

55 and recovery potential. Genetic tools are important for the identification and delineation of species (Paul D.
56 N. Hebert et al., 2003; Paul D N Hebert et al., 2003); assessment of gene-flow and sex-biased dispersal;
57 estimation of effective population size; and for determining historical changes such as range expansions
58 and contractions (DiLeo and Wagner, 2016; Shaffer et al., 2015). Such studies have historically relied on
59 small numbers of genetic markers such as simple sequence repeats (SSRs, microsatellites) (e.g. (McCracken
60 et al., 1999; Scott et al., 2001; Blouin-Demers and Gibbs, 2003; Bond et al., 2005), although next-generation
61 sequencing approaches have facilitated the identification of much larger numbers of SSRs (Castoe et al.,
62 2012). More recently, reduced representation sequencing approaches such as restriction site-associated
63 DNA sequencing (RADseq (Baird et al., 2008; Davey and Blaxter, 2010; Hohenlohe et al., 2010; Peterson et
64 al., 2012)) and genotyping-by-sequencing (GBS (Elshire et al., 2011; Narum et al., 2013)) have enabled the
65 identification and characterisation of huge numbers of single nucleotide polymorphism (SNP) markers. SSRs
66 are more variable compared to typically diallelic SNPs, but SNPs are more abundant than SSRs, and are
67 distributed more evenly across the genome, including in coding regions. SNPs can therefore provide not
68 only broader genome coverage, but also greater statistical power (Morin et al., 2004; Zimmerman et al.,
69 2020), and can differentiate between even very closely related individuals (Kleinman-Ruiz et al., 2017;
70 Roques et al., 2019), something that can be particularly important in small, isolated, inbred populations.
71 Conservation genomics studies require sources of DNA, and in the case of reptiles this can include a diverse
72 set of tissues and sampling methodologies, including both invasive and non-invasive methods. Invasive
73 sampling might include blood samples from a vein or via cardiac puncture (Brown, 2010; Eatwell et al.,
74 2014), or collection of a small amount of tissue such as a toe or tail tip, or a scale clipping (Beebee, 2008;
75 Maigret, 2019). These approaches have the disadvantages of requiring not only capture and
76 handling/restraint of the animal, but also carry with them greater ethical implications and are likely to
77 require special licensing. Non-invasive techniques are therefore to be preferred, and these can include fecal
78 sampling, the use of road kill and/or museum specimens, cloacal or buccal swabs, or shed skins (Beebee,
79 2008; Miller, 2006; Jones et al., 2008; Lanci et al., 2012; Pearson et al., 2015). While non-invasive
80 approaches are increasingly favoured, they are not without their own difficulties. Chemical preservation of
81 museum specimens can degrade DNA, and procurement of adequate samples from roadkill specimens

82 (while often high quality if found soon after death) is sporadic and unpredictable. Fecal samples inevitably
83 contain large amounts of microorganisms, and DNA often degrades quickly unless samples are rapidly
84 frozen (Jones et al., 2008). Cloacal and buccal swabbing (Beebee, 2008; Miller, 2006; Pidancier et al., 2003)
85 is dependent on locating and restraining animals with research on venomous animals carrying particular
86 risks. Shed skin samples can be collected without the need to actually locate and handle/restrain the animal
87 itself, and without associated ethical and animal welfare issues. Such risk-free approaches lend themselves
88 especially to citizen science projects, where members of the public can collect and ship samples, and
89 indeed the Amphibian and Reptile Groups of the UK (ARG UK) and Amphibian and Reptile Conservation
90 Trust (ARC) currently run a 'Reptile Slough Genebank Project' (https://www.arguk.org/get-
91 involved/projects-surveys/the-reptile-slough-genebank), which asks members of the public to send them
92 any shed skins they might find. DNA derived from shed skins of lizards and snakes has long been known to
93 be of sufficient quality and quantity for PCR-based genotyping of a small number of genetic markers
94 (Bricker et al., 1996; Fetzner Jr, 1999; Villarreal et al., 1996; Horreo et al., 2015; Tawichasri et al., 2017), and
95 so here we assess the utility of shed skin for larger-scale single nucleotide polymorphism genotyping using
96 genotyping-by-sequencing (GBS).
97
98 **Methods**
99 *DNA extraction*
100 Shed skins were collected from 61 corn snakes (*Pantherophis guttatus*) from our in-house colony, and from
101 commercial breeders and hobbyist keepers (see Supplemental Table S1 for sample details). Skins were
102 collected as soon as possible after shedding, and stored at -20°C. Those collected for us by others were
103 placed into individual paper envelopes for shipping and were stored at -20°C upon arrival in Bangor. DNA
104 was extracted from around 50mg samples of ventral scale skin using the DNeasy Blood and Tissue kit
105 (Qiagen) according to the manufacturer's protocol, with the exception of a longer (24 hour) proteinase
106 treatment at 56°C. Small DNA fragments were removed by spin-column chromatography with Chroma-
107 Spin-1000+TE columns (Clontech) following the manufacturer's protocol, and samples were quantified
108 using the Qubit dsDNA BR assay kit and Qubit Fluorometer. In some cases, it was necessary to perform

109 multiple extractions from a single sample and pool them using ethanol precipitation to obtain the desired
110 100ng of DNA. We also prepared DNA from 50mg samples of snap-frozen muscle from a further 18
111 individuals using the same procedure.
112
113 *Genotyping-by-sequencing*
114 Libraries were prepared as described by Elshire et al. (Elshire et al., 2011), at the Institute of Biological,
115 Environmental and Rural Sciences (IBERS) at Aberystwyth University. Restriction digestion was carried out
116 using the type II restriction endonuclease PstI (cut site CTGCA^G), and the resulting fragments were tagged
117 with unique barcodes of varying lengths (Supplemental Table S1). GBS libraries were pooled to an
118 equimolar concentration and single-end sequenced on one lane of Illumina HiSeq 2500.
119
120 *Bioinformatics*
121 We used the stacks (v1.44) pipeline (Catchen et al., 2011, 2013) to do a genome-guided stacks assembly
122 and call SNPs. Our pipeline started with the program 'process_radtags' and took as arguments the single
123 fastq file (-f), the list of barcodes (-b), the restriction enzyme (-e pstI), as well as the flags to clean the data
124 (-c), discard reads with low quality (-q), rescue the barcodes (-r), specify the quality encoding (-E phred33),
125 and specify how the barcodes were situated in the reads (--inline_null). Once process_radtags had finished
126 de-multiplexing and cleaning the data, we counted the remaining high-quality reads for each barcode. We
127 aligned all reads to the corn snake genome (GCA_001185365.1 (Ullate-Agote et al., 2014)) with BWA (Li and
128 Durbin, 2009) and counted the depth of coverage with 'samtools depth' (Li et al., 2009). We processed the
129 genome alignments with 'pstacks' using a minimum stack depth of 3 (-m 3), and then built the catalog with
130 'cstacks' using all individuals and allowing for 1 mismatch against the reference (-n 1), and then ran 'sstacks'
131 with all default settings. Finally, we treated all individuals as belonging to a single group and ran
132 'populations' to call variants. We included flags to keep SNPs that are present in a single population (-p 1),
133 required a minimum of 5 reads to call a stack at a locus (-m 5), kept only SNPs (--remove-indels), and export
134 the variant calls in vcf format (--vcf). We then filtered the variant file using Vcftools v0.1.15 (Danecek et al.,
135 2011) for sites with more than 60% completeness (--max_missing .6), calculated the percent missing

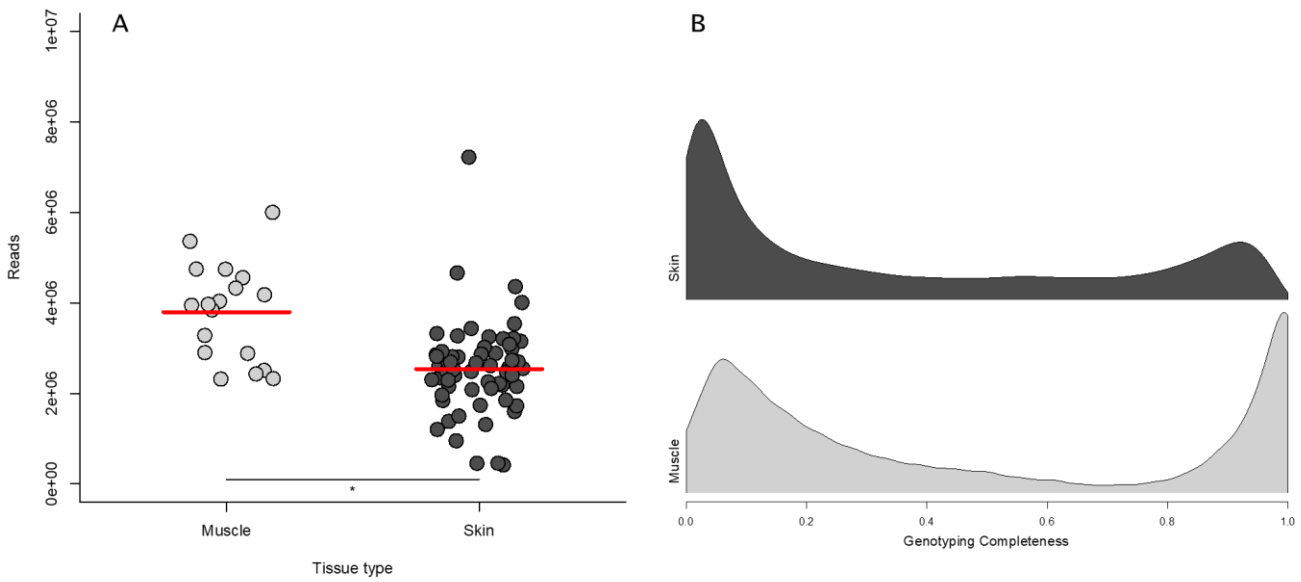
136 genotypes for each individual (--missing-indv), and the transition-transversion profiles (--TsTv-summary) for
137 both skin and muscle libraries separately, and then calculated the genome-wide Weir and Cockerham mean
138 Fst between the skin- and muscle-derived libraries (--weir-fst-pop).
139
140 *Identification of sex chromosome-specific markers*
141 We used the coverage from each individual to identify sex-linked genomic contigs as described previously
142 (Brekke et al., 2018, 2019) by first calculating each individuals' sequencing effort as the sum of aligned
143 reads for that individual. We standardised the contig-level counts by dividing by the sequencing effort of
144 each individual and multiplying by 1,000,000. We compared the mean of the coverage of females and the
145 mean coverage of males for each contig. W contigs should be present in females but not males and fulfil
146 the inequality:
147 $$Coverage_{Male} < 0.1 * Coverage_{Female} - 0.2$$
148 Unknown contigs are the not W-linked and have overall standardized coverage of less than 5:
149 $$Coverage_{Male} + Coverage_{Female} < 5$$
150 Z-linked contigs are not unknown and have twice the coverage in males as females and fulfil the inequality:
151 $$Coverage_{Male} < 1.5 * Coverage_{Female} + 0.2$$
152 All remaining contigs are annotated as autosomal. These specific cut-offs were chosen based on the
153 natural breakpoints in the plot.
154
155 *Chromosomal distribution of SNPs*
156 In late 2019, the DNA Zoo project (https://www.dnazoo.org/), released a chromosome-scale assembly of
157 the corn snake genome, comprising 119,289 contigs (contig N50 37.9kb) in 33,440 scaffolds (scaffold N50
158 147Mb), assigned to the 18 pairs of chromosomes (8 macro and 10 micro), generated using their 3D *de*
159 *novo* assembly methodology (Dudchenko et al., 2017). This assembly is based on the initial assembly of
160 Ullate-Agote et al. (Ullate-Agote et al., 2020, 2014). We determined the chromosomal distribution of our
161 SNPs by re-running the stacks pipeline as outlined above on this new chromosome-scale assembly.
162

163 **Results**

164 DNA was successfully extracted from all skin and muscle samples, with concentrations varying between
165 0.37-12ng/µl (mean 6.15ng/µl). In all cases, only a small proportion of a shed skin (typically less than
166 200mg) was needed to obtain sufficient DNA for GBS. On average there were 2,818,124 ± 1,154,007 reads
167 sequenced per individual. GBS libraries extracted from muscle tissue had more sequenced reads than
168 libraries from skin tissue (Fig. 1a, muscle mean: 3,798,510 reads, skin mean: 2,538,013 reads, Welch two
169 sample T-test, t=4.4064, df=26.183, *P*=0.0001589).

170

171



172

173 **Figure 1 (a)** Muscle-derived libraries have more reads than skin-derived libraries despite being pooled in
174 equimolar ratios during library construction (Welch two sample T-test, t=4.4064, df=26.183, *P*=0.0001589).
175 This pattern is likely caused by lower quality DNA from dried shed skins than snap-frozen muscle tissue.
176 Each sample is plotted according to its tissue of origin and the red lines denote the mean. **(b)** Density
177 distributions of the proportion of libraries for which each of the 237,466 raw SNPs are genotyped show that
178 fewer sites are genotyped in skin-derived libraries. SNPs at 0 are genotyped in no library, while SNPs at 1
179 are genotyped in all libraries. In both skin- and muscle-derived libraries there are many SNPs that are only
180 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.

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 1b, 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 2).

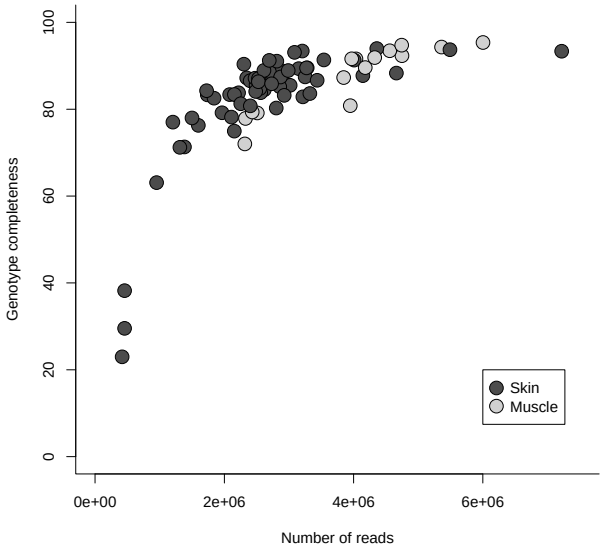


Figure 2 There is a strong relationship between the number of reads sequenced per sample and the number of missing genotypes per individual such that with deeper sequencing more variants can be called. There is a sharp cutoff at around 1,000,000 reads under which the genotype completeness drops sharply, and a consistently high genotyping rate does not occur until above 2,000,000 reads per sample. This figure shows the subset of SNPs at the >60% completeness cutoff for each library type.

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 (Figure 3, 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.

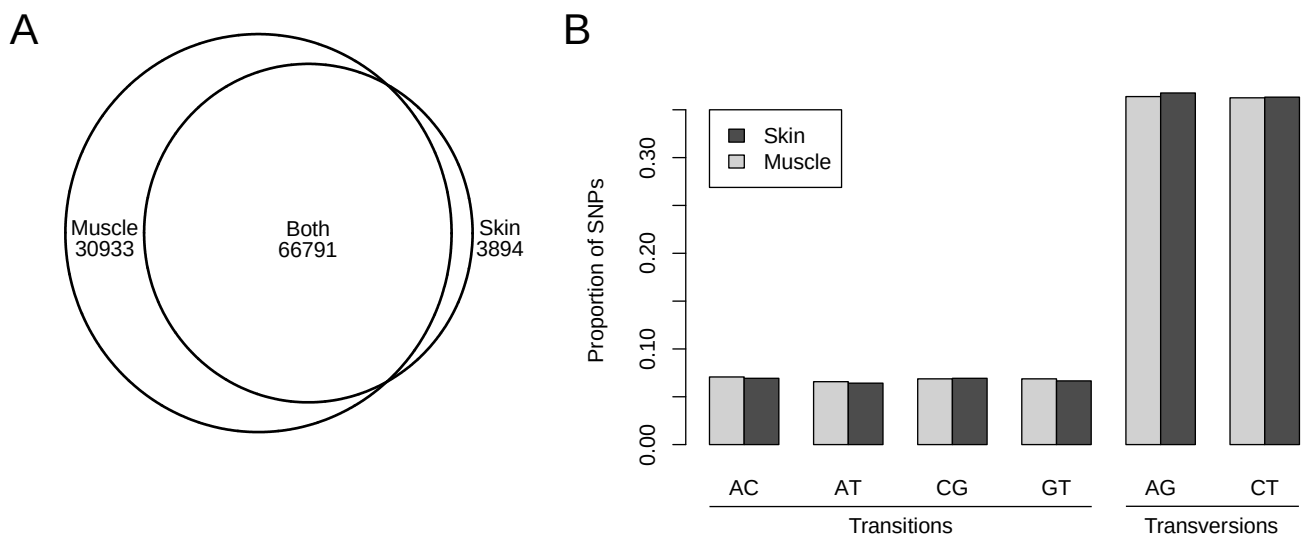


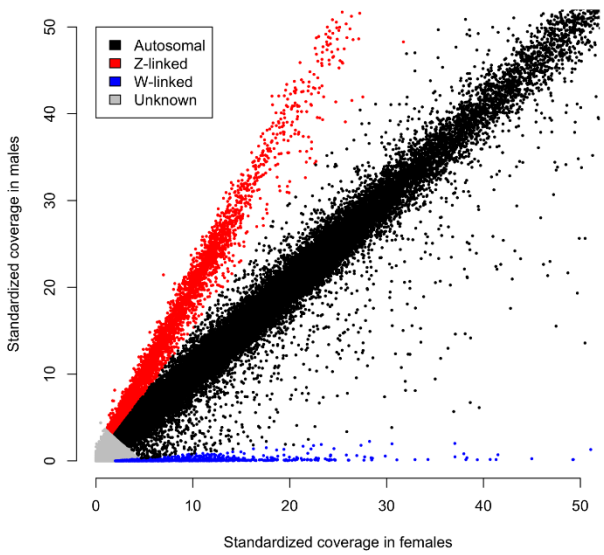
Figure 3 There are slight differences in SNP profiles between skin and muscle libraries. **(A)** After filtering each set of libraries independently for 60% completeness, most SNPs were found in both muscle and skin. **(B)** There are significant differences in the proportions of all SNP types (transitions: AC, AT, CG, and GT, and transversions: AG and CT) between skin and muscle (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

220 identifying and accounting for specific errors will be difficult and provide little benefit to the experimental
221 outcome.

222

223 We used coverage to identify the sex-linked scaffolds in the 2014 version of the corn snake genome
224 (Ullate-Agote et al., 2014) and were able to reliably annotate approximately 30% of the genome (Figure 4).
225 58,935 scaffolds (349,080,366 bases) are autosomal, 4178 scaffolds (19,511,706 bases) are Z-linked, and
226 1275 scaffolds (2,357,273 bases) are W-linked, the remaining 819,528 scaffolds (1,033,270,996 bases) had
227 too little coverage to assign (Supplemental file 1).

228



229

230 **Figure 4** Relative coverage in males and females discriminates sex-linked from autosomal scaffolds. Points
231 on the 1:1 line have equal sequencing coverage in males and females implying autosomal linkage. Z-linked
232 scaffolds (red points) and W-linked scaffolds (blue points) are also readily apparent. Scaffolds with too little
233 coverage to reliably discriminate relative coverage in males and females are shown in grey

234

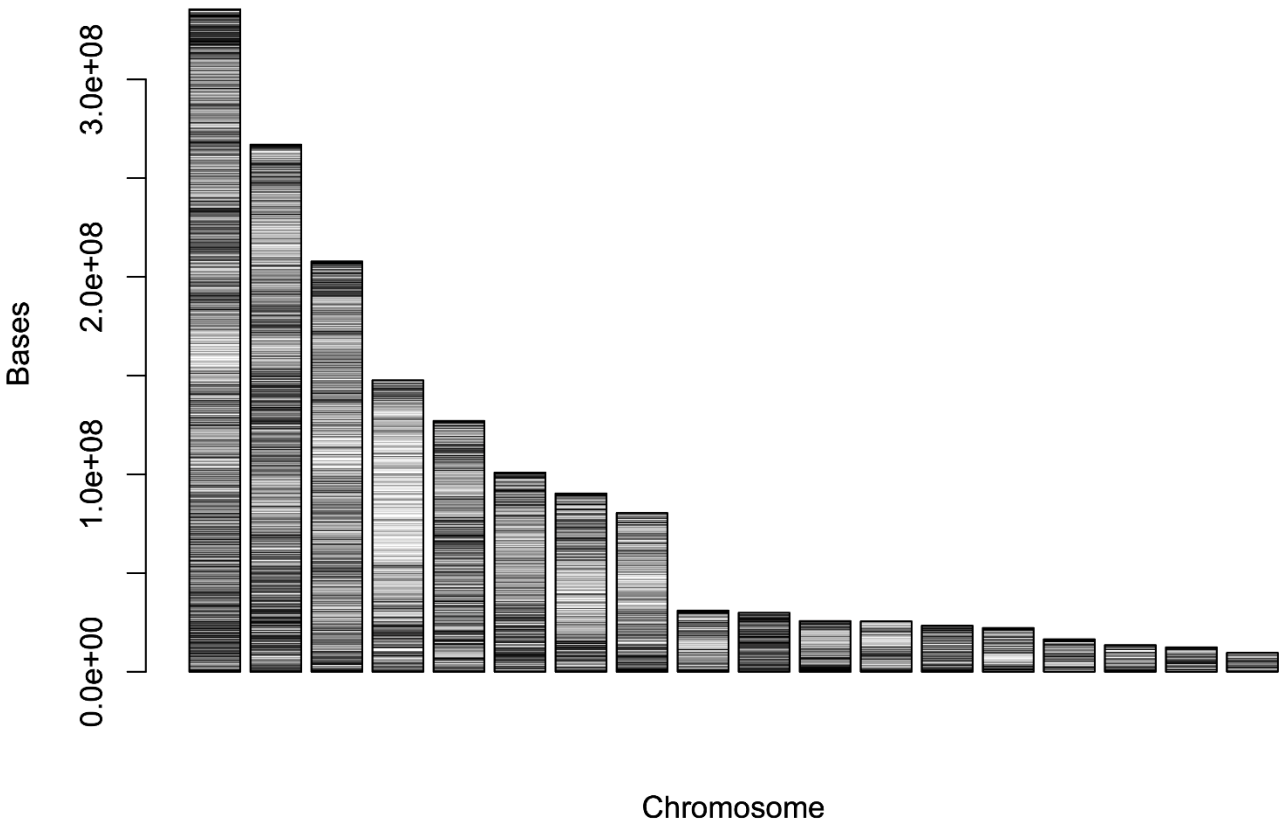
235 When we aligned the 97,638 SNPs with greater than 60% completeness to the recent chromosome-scale
236 corn snake genome assembly (Figure 5), we found that the average SNP density across all chromosomes

237 was 68.13 SNPs/Mb (69.93 SNPs/Mb across all autosomes, and 37.44 SNPs/Mb for the Z), but that there
238 were significantly more SNPs on microchromosomes (mean density 78.39 SNPs/Mb) than
239 macrochromosomes (mean density 55.30 SNPs/Mb (T-test, t = 4.3533, df = 16, *P* =0.000246), Figure 6 and
240 Supplemental Table S2).

241

242

243



244

245 **Figure 5** Distribution of SNPs in the corn snake genome. The corn snake has 8 macrochromosomes with an
246 average SNP density of 55.3/Mb, and 10 microchromosomes with an average SNP density of 78.4/Mb.
247 Assignment of chromosome 4 as the Z is based on the localisation of chromosome-specific genes
248 (Matsubara et al., 2006).

249

250 SNP density tends to scale with chromosome length, with the exception of the Z chromosome, which has
251 fewer SNPs than would be expected given its size (37.44 SNPs/Mb vs the non-Z macrochromosome average

57.85 SNPs/Mb and all-autosome average 69.93 SNPs/Mb), and the (micro)chromosome 10, which has the highest SNP density in the corn snake genome (100.51 SNPs/Mb) (Figure 7).

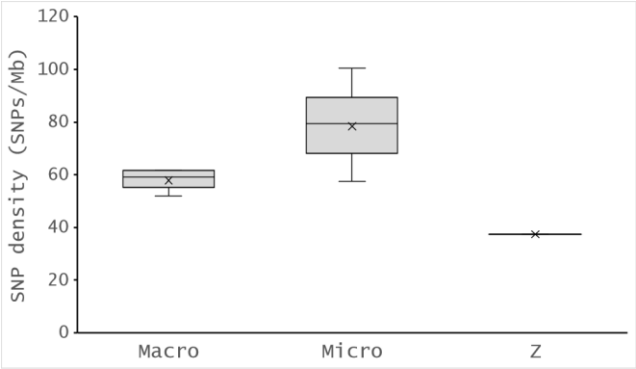


Figure 6 Microchromosomes have significantly more SNPs assigned to them than non-Z macrochromosomes (mean 78.39 SNPs/Mb vs 57.85). The Z chromosome has the lowest SNP density of any corn snake chromosome in this study (37.44 SNPs/Mb)

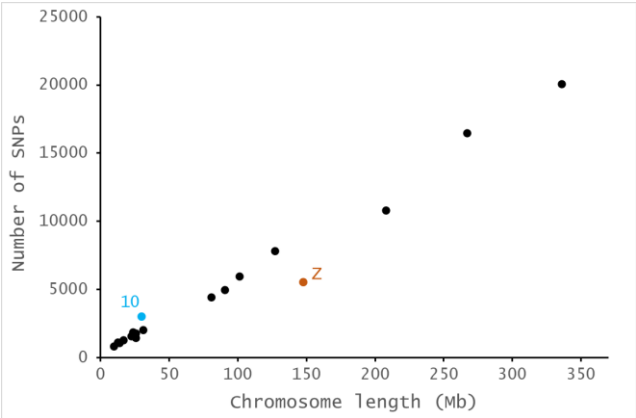


Figure 7 SNP density scales with chromosome length, with the exception of the Z chromosome (orange), which has fewer than expected, and microchromosome 10 (blue), which has the highest SNP density in the corn snake genome (100.51 SNPs/Mb, see Supplemental Table S2).

Discussion

268 DNA extracted from shed skins is suitable for use with reduced-representation sequencing
269 approaches such as GBS with some considerations. We compared GBS libraries built from skin-extracted
270 DNA with libraries from muscle-extracted DNA. The muscle samples were immediately snap-frozen and are
271 thus a source of high-quality DNA while the skins were collected by pet snake owners across the country,
272 dried, and shipped at ambient temperature through the post and thus subject to a variety of mechanisms
273 of DNA degradation. In this way they more accurately reflect the type of material that may be collected
274 from captive specimens, or what may be obtained via citizen science projects such as the 'Reptile Slough
275 Genebank Project'. We found significant differences between the muscle- and skin-derived DNA that likely
276 stem from DNA damage in the original samples. These differences include the number of reads sequenced
277 (Figure 1a) and SNPs identified (Figure 3). However, dealing with the differences between the two library
278 types is surmountable with some forethought toward experimental design.
279 The number of SNPs identified in a GBS experiment depends strongly on the sequencing depth (Figure 2).
280 Thus, sequencing skin samples more deeply than samples from fresh tissue will help assure that a sufficient
281 number of real SNPs can be identified. Estimating the necessary coverage in a GBS experiment is difficult as
282 it depends on the frequency of cut-sites in the genome (often unknown in non-model species), the
283 heterozygosity present in the sample population (often unknown until at least after the first round of
284 sequencing), the specific cut-offs used to filter variant sites, and the type of experiment (population
285 genetics may require different number of SNPs than parentage analyses etc). For this experiment the SNP
286 discovery curve plateaued at around 3,000,000 reads per individual (Figure 2), and many of the skin-derived
287 libraries had far fewer genotyped SNPs due to low read coverage (Figure 1b and 2). As such, we suggest
288 that future researchers plan on sequencing skin and other possibly damaged samples more deeply in order
289 to identify a robust set of variants.
290 Some problems, especially those relating to DNA damage, cannot be dealt with simply by sequencing more
291 deeply. For instance, if Cytosine deamination into Uracil is a common issue in the sample (as is often the
292 case in ancient DNA studies (Hofreiter et al., 2001), more sequencing will not remove those errors and a
293 more sophisticated approach is needed. The signal of errors in the skin libraries is very slight suggesting
294 that identifying and dealing with these errors would be difficult. Fortunately, such labour-intensive error

295 cleaning will likely provide little benefit for three reasons. First, the vast majority of the SNPs identified in
296 the skin samples are also found in the muscle samples (Figure 3a) and this is true despite the excess of skin-
297 derived libraries (61 vs 18 muscle-derived) in which to discover SNPs. Even if the 3,894 skin-specific SNPs
298 are not biologically real, they are so few (only 3.8% of all SNPs) that that identifying them in the absence of
299 other high-quality libraries will be exceedingly difficult and more importantly, removing them will have little
300 effect on the overarching results. Secondly, while the skin and muscle may have slightly different SNP
301 profiles (Figure 3b), the effect is so slight that filtering options such as removing all C/T sites or ignoring all
302 transversions are not merited. Finally, if there was a systematic damage pressure, it should alter the allele
303 frequencies in the damaged samples. But the Fst between the skin and muscle samples is very low (genome
304 wide mean Fst is 0.0686). While this metric can only be calculated for the 66,791 shared SNPs, it shows that
305 there are not striking allele frequency differences between skin- and muscle-derived libraries which is
306 evidence that the degradation is not a major concern. Such a low Fst result is additionally convincing given
307 that the individuals from whom muscle was used are all close relatives in our colony while the snakes
308 whose skins were used originate from breeders across the UK. This confounding population structure
309 should artificially increase our estimate of Fst, and the Fst due solely to DNA damage in the skin is likely
310 much smaller. In sum, while DNA damage is apparent in our skin samples, the effects are slight and not
311 likely to impact any biological results.
312 The SNPs we identified are not distributed evenly across the corn snake genome, although SNP density
313 does generally correlate with chromosome length (Figure 7). Microchromosmes have significantly higher
314 SNP density than macrochromosomes (Figure 6), which should perhaps not be too surprisingly given the
315 well-known differences in base composition between macro and microchromosomes (Srikulnath et al.,
316 2021). Indeed, in the corn snake, macrochromosomes have an average GC% of 39.69% and
317 microchromosomes 44.79% (Supplemental Table S2). The PstI enzyme we used in preparation of our GBS
318 libraries has a GC-rich cut site (CTGCA^G) and so is likely to cut more often on microchromosomes. Future
319 researchers may therefore wish to consider the GC-richness of microchromosomes when choosing
320 restriction enzymes for reptile GBS experiments.

321 Leveraging the amateur herpetological community and sourcing skins from the public has the dual benefit
322 of engaging citizen scientists and the possibility of rapidly collecting extremely large sample sizes. Pet
323 snakes in general, and corn snakes in particular, have a variety of colour and pattern morphs which may
324 prove incredibly powerful for understanding the genetic basis for colouration (Ullate-Agote et al., 2020,
325 2014). We have shown that a simple crowd-sourced collection technique (mailing shed snake skins) can
326 provide samples containing DNA of sufficient quality for reduced representation sequencing. This finding
327 opens up the possibility of doing association studies on patterning and colouration from a wealth of
328 samples in a long-lived and low-fecundity species, although more research is required into the impact of
329 time since shedding, and UV exposure (especially for field-collected sheds) on DNA quality.
330

331 **Acknowledgements**
332 We wish to thank Adam Clarke and numerous anonymous donors for both financial support and donation
333 of shed skins. We also wish to thank the School of Natural Science technical team, and especially Rhys
334 Morgan, for assistance with animal care, and Adam Hargreaves for useful discussions.
335

336 **Funding**
337 The authors declare that no funds, grants, or other support were received during the preparation of this
338 manuscript.
339

340 **Competing Interests**
341 The authors have no relevant financial or non-financial interests to disclose.
342

343 **Data Accessibility**
344 The datasets generated during and/or analysed during the current study are available in the European
345 Nucleotide Archive (ENA) under project PRJEB32869.
346

347 **Author contributions**

348 JFM devised the study. LS and MJH performed experiments, and TDB and JFM analysed the data and wrote
349 the paper.
350
351 **References**

352 Anon (n.d.) *The IUCN Red List of Threatened Species*. Available from: https://www.iucnredlist.org/en
353 (Accessed 21 July 2022).

354 Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A.,
355 Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers.
356 *PLOS One* 3 (10), e3376. https://doi.org/10.1371/journal.pone.0003376

357 Beebee, T. J. C. (2008). Buccal swabbing as a source of DNA from squamate reptiles. *Conservation Genetics,*
358 9 (4), 1087–1088. https://doi.org/10.1007/s10592-007-9464-2

359 Blouin-Demers, G. & Gibbs, H. L. (2003). Isolation and characterization of microsatellite loci in the black rat
360 snake (*Elaphe obsoleta*). *Molecular Ecology Notes,* 3 (1), 98–99. https://doi.org/10.1046/j.1471-
361 8286.2003.00363.x

362 Bond, J.M., Porteous, R., Hughes, S., Mogg, R.J., Gardner, M.G., Reading, C.J. (2005). Polymorphic
363 microsatellite markers, isolated using a simple enrichment procedure, in the threatened smooth
364 snake (*Coronella austriaca*). *Molecular Ecology Notes,* 5 (1), 42–44. https://doi.org/10.1111/j.1471-
365 8286.2004.00824.x

366 Brekke, T.D., Steele, K.A., Mulley, J.F. (2019). A high-density genetic map and molecular sex-typing assay for
367 gerbils. *Mammalian Genome,* 30 (3–4), 63–70. https://doi.org/10.1007/s00335-019-09799-z

368 Brekke, T.D., Supriya, S., Denver, M.G., Thom, A., Steele, K.A., Mulley, J.F. (2018). Inbred or Outbred?
369 Genetic diversity in laboratory rodent colonies. *G3: Genes, Genomes, Genetics,* 8 (2), 679–686.
370 https://doi.org/10.1534/g3.117.300495

Bricker, J., Bushar, L.M., Reinert, H.K., Gelbert, L. (1996). Purification of high quality DNA from shed skins. *Herpetological Review,* 27 (3), 133.

Brown, C. (2010). Cardiac blood sample collection from snakes. *Lab Animal*, 39 (7), 208–209. https://doi.org/10.1038/laban0710-208

Castoe, T.A., Poole, A.W., Koning, A.P.J. de, Jones, K.L., Tomback, D.F., Oyler-McCance, S.J., Fike, J.A., Lance, S.L., Streicher, J.W., Smith, E.N., Pollock, D.D. (2012). Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLOS One*, 7 (2), e30953. https://doi.org/10.1371/journal.pone.0030953

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology,* 22 (11), 3124–3140. https://doi.org/10.1111/mec.12354

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J.H. (2011). Stacks: building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics,* 1 (3), 171–182. https://doi.org/10.1534/g3.111.000240

Cox N, Young BE, Bowles P, Fernandez M, Marin J, Rapacciuolo G, et al. (2022). A global reptile assessment highlights shared conservation needs of tetrapods. *Nature,* 605 (7909), 285–290. https://doi.org/10.1038/s41586-022-04664-7

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics,* 27 (15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davey, J. W. & Blaxter, M. L. (2010). RADSeq: next-generation population genetics. *Briefings in Functional Genomics,* 9 (5–6), 416–423. https://doi.org/10.1093/bfgp/elq031

DiLeo, M. F. & Wagner, H. H. (2016). A landscape ecologist's agenda for landscape genetics. *Current Landscape Ecology Reports,* 1 (3), 115–126. https://doi.org/10.1007/s40823-016-0013-x

394 Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I.,
395 Lander, E.S., Aiden, A.P., Aiden, E.L. (2017). De novo assembly of the *Aedes aegypti* genome using
396 Hi-C yields chromosome-length scaffolds. *Science,* 356 (6333), 92–95.
397 https://doi.org/10.1126/science.aal3327

398 Eatwell, K., Hedley, J., Barron, R. (2014). Reptile haematology and biochemistry. *In Practice,* 36 (1), 34–42.
399 https://doi.org/10.1136/inp.f7488

400 Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E. (2011). A robust,
401 simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS One,* 6 (5),
402 e19379. https://doi.org/10.1371/journal.pone.0019379

403 Fetzner Jr, J. W. (1999). Extracting high-quality DNA from shed reptile skins: a simplified method.
404 *Biotechniques,* 26 (6), 1052–1054. https://doi.org/10.2144/99266bm09

405 Hebert P., Ratnasingham S., de Waard J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1
406 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences,*
407 270 (Suppl 1), S96–S99. https://doi.org/10.1098/rsbl.2003.0025

408 Hebert P., Cywinska A., Ball S. L., deWaard J. R. (2003). Biological identifications through DNA barcodes.
409 *Proceedings of the Royal Society B: Biological Sciences*, 270 (1512), 313–321.
410 https://doi.org/10.1098/rspb.2002.2218

411 Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., Pääbo, S. (2001). Ancient DNA. *Nature Reviews Genetics,* 2
412 (5), 353–359. https://doi.org/10.1038/35072071

413 Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A. (2010). Population
414 genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLOS Genetics,*
415 6 (2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

416 Horreo, J.L., Peláez, M.L., Fitze, P.S. (2015). Skin sheds as a useful DNA source for lizard conservation.
417 *Phyllomedusa: Journal of Herpetology,* 14 (1), 73–77. https://doi.org/10.11606/issn.2316-
418 9079.v14i1p73-77

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

421 Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B., Fernández, J.,
422 Godoy, J.A. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly
423 endangered Iberian lynx. *BMC Genomics,* 18 (1), 556. https://doi.org/10.1186/s12864-017-3946-5

424 Lanci, A.K.J., Roden, S.E., Bowman, A., LaCasella, E.L., Frey, A., Dutton, P.H. (2012). Evaluating buccal and
425 cloacal swabs for ease of collection and use in genetic analyses of marine turtles. *Chelonian*
426 *Conservation and Biology,* 11 (1), 144–148. https://doi.org/10.2744/CCB-0950.1

427 Li, H., Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25 (16), 2078–
428 2079. https://doi.org/10.1093/bioinformatics/btp352

429 Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform.
430 *Bioinformatics*, 25 (14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

431 Maigret, T. A. (2019). Snake scale clips as a source of high quality DNA suitable for RAD sequencing.
432 *Conservation Genetics Resources*, 11 (4), 373–375. https://doi.org/10.1007/s12686-018-1019-y

433 Matsubara, K., Tarui, H., Toriba, M., Yamada, K., Nishida-Umehara, C., Agata, K., Matsuda, Y. (2006).
434 Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise
435 differentiation of snake sex chromosomes. *Proceedings of the National Academy of Sciences*, 103
436 (48), 18190–18195. https://doi.org/10.1073/pnas.060527410

McCracken G, Burghardt G, Houts S. (1999). Microsatellite markers and multiple paternity in the garter snake *Thamnophis sirtalis*. *Molecular Ecology,* 8 (9), 1475–1479. https://doi.org/10.1046/j.1365-294x.1999.00720.x

Miller, H. C. (2006). Cloacal and buccal swabs are a reliable source of DNA for microsatellite genotyping of reptiles. *Conservation Genetics,* 7 (6), 1001–1003. https://doi.org/10.1007/s10592-006-9120-2

Morin, P.A., Luikart, G., Wayne, R.K., et al. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, 19 (4), 208–216. https://doi.org/10.1016/j.tree.2004.01.009

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R., Hohenlohe, P.A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22 (11), 2841–2847. https://doi.org/10.1111/mec.12350

Pearson, S.K., Tobe, S.S., Fusco, D.A., Bull, C.M., Gardner, M.G. (2015). Piles of scats for piles of DNA: deriving DNA of lizards from their faeces. *Australian Journal of Zoology,* 62 (6), 507–514. https://doi.org/10.1071/ZO14059

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E. (2012). Double digest RADseq: An inexpensive method for *de Novo* SNP discovery and genotyping in model and non-model species. *PLOS One,* 7 (5), e37135. https://doi.org/10.1371/journal.pone.0037135

Pidancier, N., Miquel, C., Miaud, C. (2003). Buccal swabs as a non destructive tissue sampling method for DNA analysis in amphibians. *Herpetological Journal*, 13 (4), 175–178.

Roques, S., Chancerel, E., Boury, C., Pierre, M., Acolas, M. L. (2019). From microsatellites to single nucleotide polymorphisms for the genetic monitoring of a critically endangered sturgeon. *Ecology and Evolution,* 9 (12), 7017–7029. https://doi.org/10.1002/ece3.5268

Scott, I.A.W., Hayes, C.M., Keogh, J.S., Webb, J.K. (2001). Isolation and characterization of novel microsatellite markers from the Australian tiger snakes (Elapidae: *Notechis*) and amplification in the

closely related genus *Hoplocephalus*. *Molecular Ecology Notes*, 1 (3), 117–119. https://doi.org/10.1046/j.1471-8278.2001.00037.x

Shaffer HB, Gidiş M, McCartney-Melstad E, Neal KM, Oyamaguchi HM, Tellez M, et al. (2015). Conservation genetics and genomics of amphibians and reptiles. *Annual Review of Animal Biosciences*, 3113–138. https://doi.org/10.1146/annurev-animal-022114-110920

Srikulnath, K., Ahmad, S.F., Singchat, W., Panthum, T. (2021). Why do some vertebrates have microchromosomes? *Cells*, 10 (9), 2182. https://doi.org/10.3390/cells10092182

Tawichasri, P., Laopichienpong, N., Chanhome, L., Phatcharakullawarawat, R., Singchat, W., Koomgun, T., Prasongmaneerut, T., Rerkamnuaychoke, W., Sillapaprayoon, S., Muangmai, N. (2017). Using blood and non-invasive shed skin samples to identify sex of caenophidian snakes based on multiplex PCR assay. *Zoologischer Anzeiger,* 2716–14. https://doi.org/10.1016/j.jcz.2017.11.003

Ullate-Agote, A., Burgelin, I., Debry, A., Langrez, C., Montange, F., Peraldi, R., Daraspe, J., Kaessmann, H., Milinkovitch, M.C., 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), 26307–26317. https://doi.org/10.1073/pnas.200372411

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. doi: 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. https://doi.org/10.1093/oxfordjournals.jhered.a022973

Zimmerman, S.J., Aldridge, C.L., Oyler-McCance, S.J. (2020). An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC Genomics*, 21 (1), 382. https://doi.org/10.1186/s12864-020-06783-9

Supplementary tables

Title

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

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

Supplementary Table S1. Sample details.

Supplementary Table S2. Distribution of SNPs in the corn snake genome.

514 **Supplementary Table S1.** Sample details.

Sample ID	Barcode	Sex	Sample type
PC13	GAATACA	Female	Shed skin
PC22	CGCTGA	Female	Shed skin
PC25	CACCATGAA	Female	Shed skin
PC28	CACATGA	Female	Shed skin
JM37	AAGTCA	Female	Shed skin
JM38	GAACGACCT	Female	Shed skin
PC36	TAGACCA	Female	Shed skin
PC09	TTGCTTCA	Female	Shed skin
PC10	AACT	Female	Shed skin
PC08	CCACG	Female	Shed skin
PC17	TGATCA	Female	Shed skin
PC35	ACCTCATG	Female	Shed skin
JM3	TTCCGCA	Female	Shed skin
PC02	TTCGAACA	Female	Shed skin
PC20	GGTGAGTCT	Female	Shed skin
PC29	ACGAG	Female	Shed skin
PC43	GGAACAAT	Female	Shed skin
PC05	TGTCACA	Female	Shed skin
PC12	GAGAACAT	Female	Shed skin
PC16	AAGATGCA	Female	Shed skin
PC18	AACGAACAT	Female	Shed skin
PC19	ATTG	Female	Shed skin
PC27	CGTGAGGT	Female	Shed skin
PC33	GAGCG	Female	Shed skin
PC41	CAGAT	Female	Shed skin
PC07	TTGCTTACA	Female	Shed skin
PC06	TATAA	Female	Shed skin
JM6	AGTGGCGTT	Female	Shed skin
PC03	GGAACAAGT	Female	Shed skin
PC34	CTGCCA	Female	Shed skin
PC40	GATCCA	Male	Shed skin
JM4	CACTAG	Male	Shed skin
JM8	AGAATCA	Male	Shed skin
PC23	CCAG	Male	Shed skin
PC26	CTTGCAA	Male	Shed skin
PC11	TTGTTCCTA	Male	Shed skin
PC15	GAATACCA	Male	Shed skin
PC39	CTTCTGTA	Male	Shed skin
JM30	CCTGTCA	Male	Shed skin
JM30	TCTAGCA	Male	Shed skin
JM35	ACATCG	Male	Shed skin
JM35	ACATA	Male	Shed skin
JM9	GGTA	Male	Shed skin
PC44	CCAACCTGA	Male	Shed skin

PC14	TCCGACA	Male	Shed skin
PC21	CTTACGCA	Male	Shed skin
PC24	AACCA	Male	Shed skin
PC30	ACTGCA	Male	Shed skin
PC37	ACCGTCCA	Male	Shed skin
PC42	TTGA	Male	Shed skin
JM80	AACGTG	Male	Shed skin
JM81	GGCCTCA	Male	Shed skin
JM82	GCTGGACA	Male	Shed skin
JM83	TAGCGACA	Male	Shed skin
JM84	AGTCG	Male	Shed skin
JM85	GCCAA	Male	Shed skin
PC04	GGACTA	Male	Shed skin
PC31	TTCTA	Male	Shed skin
PC01	GAAGA	Male	Shed skin
PC32	GGATCAA	Male	Shed skin
PC38	TCGCAA	Male	Shed skin
JM1	GTGGCCA	Male	Shed skin
JM112	TTCGCA	Unknown - juvenile	Muscle sample
JM100	CCTAAG	Unknown - juvenile	Muscle sample
JM101	TTACGA	Unknown - juvenile	Muscle sample
JM102	CATCGCA	Unknown - juvenile	Muscle sample
JM104	CCTACACCG	Unknown - juvenile	Muscle sample
JM105	TCTGG	Unknown - juvenile	Muscle sample
JM106	ACCATGACA	Unknown - juvenile	Muscle sample
JM108	GAACAGAAT	Unknown - juvenile	Muscle sample
JM109	CGGT	Unknown - juvenile	Muscle sample
JM110	CGTACA	Unknown - juvenile	Muscle sample
JM111	AACAGAATA	Unknown - juvenile	Muscle sample
JM113	GTTCA	Unknown - juvenile	Muscle sample
JM91	CCACCTCGT	Unknown - juvenile	Muscle sample
JM92	CCATGCCA	Unknown - juvenile	Shed skin
JM94	CTCAG	Unknown - juvenile	Muscle sample
JM96	ACATGCA	Unknown - juvenile	Muscle sample
JM97	TAATG	Unknown - juvenile	Muscle sample
JM98	TGCG	Unknown - juvenile	Muscle sample
JM99	ACCTACCG	Unknown - juvenile	Muscle sample

526 **Supplementary Table S2.** Distribution of SNPs in the corn snake genome.

	Chromosome	Number of SNPs	Chromosome length (Mb)	SNP density (SNPs per Mb)	GC%
Macro	1	20070	335.6	59.80	39.21
	2	16490	267.0	61.77	39.64
	3	10794	207.8	51.94	39.12
	4 (Z)	5526	147.6	37.44	40.02
	5	7831	127.0	61.65	39.53
	6	5978	100.9	59.23	39.58
	7	4977	90.3	55.11	40.40
	8	4454	80.3	55.45	40.00
Micro	9	2022	30.9	65.34	42.97
	10	3000	29.8	100.51	44.08
	11	1772	25.7	68.94	42.95
	12	1464	25.5	57.45	43.08
	13	1862	23.3	80.00	43.72
	14	1570	22.2	70.78	44.82
	15	1295	16.4	78.76	45.93
	16	1093	13.4	81.30	46.23
	17	1138	12.3	92.29	47.00
	18	847	9.6	88.55	47.09

527

528 Summary statistics:

	SNP density		
	Mean	**Median**	**Mean GC%**
All chromosomes	68.13	63.55	42.52
All macrochromosomes	55.30	57.34	39.69
Macrochromosomes (no Z)	57.85	59.23	39.64
Microchromosomes	78.39	79.38	44.79
All autosomes	69.93	65.34	42.67

529