TWYNI AR SYMUD | DYNAMIC DUNESCAPES

CRONFA DREFTADAETH HERITAGE FUND

Cyfoeth Naturiol Cymru Natural Resources Wales

Dynamic Dunescapes

Genetic ancestry of introduced populations of sand lizard (*Lacerta agilis*) in Wales

July 2022

NATURAL ENGLAND

Plantlife

National Trust

Cyfoeth Naturiol Cymru Natural Resources Wales

The Wildlife Trusts

Genetic ancestry of introduced populations of sand lizard (*Lacerta agilis*) in Wales

Report No: **NRW Evidence Report No 596**

Authors Name: [1,2]**John Benjamin Owens,** [3]**John W. Wilkinson,** [1]**Wolfgang Wüster,** [1]**Axel Barlow,** [1]**Alexander Papadopulos**

[1]Molecular Ecology and Evolution at Bangor, School of Natural Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, Wales.

[2]Captive & Field Herpetology Ltd, 13 Hirfron, Llaingoch, Holyhead, Anglesey, LL65 1YU, Wales.

[3]Amphibian and Reptile Conservation Trust, 744 Christchurch Road, Boscombe, Bournemouth, Dorset BH7 6BZ, England.

About Natural Resources Wales

Natural Resources Wales' purpose is to pursue sustainable management of natural resources. This means looking after air, land, water, wildlife, plants and soil to improve Wales' well-being, and provide a better future for everyone.

Evidence at Natural Resources Wales

Natural Resources Wales is an evidence-based organisation. We seek to ensure that our strategy, decisions, operations and advice to Welsh Government and others are underpinned by sound and quality-assured evidence. We recognise that it is critically important to have a good understanding of our changing environment.

We will realise this vision by:

- Maintaining and developing the technical specialist skills of our staff;
- Securing our data and information;
- Having a well-resourced proactive programme of evidence work;
- Continuing to review and add to our evidence to ensure it is fit for the challenges facing us; and
- Communicating our evidence in an open and transparent way.

This Evidence Report series serves as a record of work carried out or commissioned by Natural Resources Wales. It also helps us to share and promote use of our evidence by others and develop future collaborations. However, the views and recommendations presented in this report are not necessarily those of NRW and should, therefore, not be attributed to NRW.

Report series:	**NRW Evidence Report**
Report number:	**No 596**
Publication date:	**March 2022**
Contract number:	**NRW Evidence Report No 596**
Contractor:	**University of Bangor**
Contract Manager:	**Emma Brown**
Title:	**Genetic ancestry of introduced populations of sand lizard (*Lacerta agilis*) in Wales**
Author(s):	**John Benjamin Owens, John W. Wilkinson, Wolfgang Wüster, Axel Barlow, Alexander Papadopulos**
Technical Editor:	**Matthew Ellis & Julie Creer**
Quality assurance:	**Tier 2 sign off; Emma Brown, Julie Creer, Sam Dyer & Matthew Ellis**
Peer Reviewer(s):	**Julie Creer, Sam Dyer & Matthew Ellis**
Approved By:	**Nick Thomas & Emma Brown**
Restrictions:	**None**

Distribution List (core)

NRW Library, Bangor	2
National Library of Wales	1
British Library	1
Welsh Government Library	1
Scottish Natural Heritage Library	1
Natural England Library (Electronic Only)	1

Distribution List (others)

Amphibian and Reptile Conservation Trust, Bournemouth	3
Matthew Ellis, NRW, Wrexham	1
Chris Worker, Welsh Government, Aberystwyth	1
Wildlife Trust of South and West Wales	2

Acknowledgements

The authors would like to express gratitude to Liam Russell, Stuart Graham, Pete Hill, Mark Barber and George Mee for samples; Andrea Pozzi for laboratory assistance; James Hicks, Maya Master, Mark Barber, Emma Brown, Chris Worker and Matt Ellis for comments, guidance and other assistance during this investigation. We acknowledge the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via Welsh Government.

The report has been produced for the Dynamic Dunescapes project with support from the National Lottery Heritage Fund and match funding provided by NRW.

Recommended Citation for this Volume:

Owens JB, Wilkinson JW, Wüster W, Barlow A, Papadopulos A. 2022. Genetic ancestry of introduced populations of sand lizard (*Lacerta agilis*) in Wales. NRW Evidence Report. Report No: 596, 30pp, Natural Resources Wales, Bangor.

Image: Dune race sand lizard (*Lacerta agilis*) from the Sefton Coast, Merseyside. © John Benjamin Owens

Contents

List of Figures

disregarding invariant positions, and are therefore much higher than the actual genome divergence will be.

Figure 6. Mean proportion of heterozygous sites (y axis) for the investigated sand lizards (x axis). Values are based on analysis of Chromosome 1 (~130 megabases). Results for the Gower individuals are shown in red.

Figure 7. Runs of homozygosity in introduced sand lizard populations. The x axis represents chromosome 1 with the position indicated in megabases (of around 130 megabases total length). Individual bars show the proportions of heterozygous position for a 1 megabase window (Y axes, all scaled between 0–0.015). Runs of homozygosity are highlighted in red. The length of the sampled chromosome differs slightly between individuals due to variable coverage and the effects of filtering, but this represents a maximum of only 8 windows between individuals and can be generally ignored.

Figure 8. Runs of homozygosity in natural sand lizard populations. For details see Figure 7 legend.

Plate 1. Male sand lizards. 1) Specimen from Sefton Coast, Merseyside. 2) Specimen from the Gower population. Note the typical pattern of the dune race, with largely uniform green sides in males, common to both populations. © Pete Hill

List of Tables

Executive Summary

Despite its wide Eurasian range, the sand lizard (*Lacerta agilis*) is the rarest and most threatened lizard species within the UK, where it is restricted to sandy heathlands and sand dunes in southern and north-western England and North Wales. The species has suffered extensive declines and extinctions across the UK due to development of its coastal and heathland habitats. As a result, its populations in the UK have become highly fragmented, and the species had become extinct in parts of its range, including Wales. Ex-situ captive breeding and reintroduction programmes have been successful in restoring declining and extinct populations in England and Wales. In North Wales, where the species had been extinct for over 50 years, captive-bred animals from Merseyside have been used successfully to restore several coastal sand dune populations.

However, over the last two decades, sand lizards have also been reported in areas where they were not previously known and that were not included in planned, official

reintroduction programmes. This indicates the presence of “unofficial” reintroductions, in particular at Newborough Warren and Aberffraw, Anglesey.

In 2018, an additional, apparently thriving population of sand lizards was discovered on the Gower Peninsula, South Wales. The location is separated from any current or historic sand lizard records by extensive areas lacking suitable habitats. This raises questions on the origin of this population. Plausible hypotheses are that this population could represent:

i. a hitherto overlooked native population

ii. an “unofficial” release of British sand lizard stock by private breeders, or

iii. a colony descended from escaped/released pet lizards from elsewhere in their extensive European and Asian range.

The aim of this investigation was to use modern genomic approaches to assess the ancestry of the Gower sand lizards to inform conservation decisions on this population, and a preliminary assessment of its genetic health.

We used medium coverage whole genome sequencing to assess the genetic ancestry of unofficial introductions of sand lizards in Wales. This state-of-the-art approach provides the best option to achieve the project aims as it provides exceptional analytical resolution of population ancestry even from small numbers of sampled individuals. We sequenced sand lizard buccal swab samples representing the newly discovered Gower population as well as populations in England (Dorset, Surrey, Merseyside), North Wales (Harlech, Aberffraw) and the Netherlands, and used published data from a Danish individual, in order to assess the origin of the Gower sand lizards. We assessed the relationships among samples using principal components analysis (PCA), phylogenetic analysis, and comparison of pairwise genetic distances. We aligned mitochondrial sequences from our samples with published *L. agilis* sequences from across their distribution to place them in the broader context of sand lizard phylogeography. Finally, we measured heterozygosity and tested for runs of homozygosity in chromosome 1 to assess the genetic health of the Gower population.

Key results:

- mitochondrial DNA analysis: a phylogenetic analysis of 908 b.p. of the cytochrome *b* gene confirmed that the Gower specimens share the same haplotype as all other UK samples, supporting a UK origin of the founder stock of the Gower population.
- PCA analysis of 395,413,986 single nucleotide polymorphisms (SNPs) in the genome sequences grouped the Gower sand lizards with North Wales and Merseyside dune populations, separate from populations from southern England or continental Europe.
- Phylogenetic analysis of the genome data identified the “dune race” sand lizard from Merseyside as a cohesive group, and groups the Gower specimens as part of that group, without clear separation from them.
- Pairwise genetic distances show that all populations of the "dune race” sand lizards, including the Gower population, can be considered as a single genetic entity
- Analysis of genetic health showed that all sand lizards investigated have high levels of heterozygosity. No runs of homozygosity were observed in the Gower individuals but were apparent in several other populations indicating inbreeding. Thus, the genetic health of the Gower population currently seems to be high within the context of the UK as a whole.

Our analyses demonstrate that sand lizards on the Gower Peninsula are unambiguously of UK origin, and genetically indistinguishable from the populations in North Wales and Merseyside (all ultimately descended from Merseyside stock). They exclude the possibility of a continental European or admixed origin. The lack of genomic differentiation of the Gower population suggests a very recent common ancestry with the north-western dune lizards, and the lack of suitable intervening habitat argues for a recent “unofficial” release of north-western dune lizards as the origin of the Gower population. The population appears to be in good genetic health.

The sand lizard populations on the Gower represent an outlying population of pure north-western “dune race” sand lizards. While their presence on the Gower Peninsula is probably due to an “unofficial” reintroduction, the presence of this apparently thriving and genetically healthy population contributes positively to the overall conservation status of the species, and the north-western dune race in particular, in Wales and the UK.

The genomic resources generated during this study will serve as a valuable baseline for the inexpensive application of whole genome sequencing approaches to any further unexpected *Lacerta agilis* populations in the UK.

Introduction

Background & Rationale

The sand lizard (*Lacerta agilis*) is one of the rarest reptiles in the UK and is Wales’ rarest reptile, due to its specialised habitat requirements. They are listed as a priority species under the UK Biodiversity action plan (BAP), Wildlife and Countryside Act (1981) and as a European protected species under the Conservation of Habitats and Species Regulations (2010) (Soorae, 2011). Confined to mobile sand dune systems and sandy heathland habitats in Wales/north-west England and southern England, respectively, sand lizards occupy isolated and fragmented patches of suitable habitat along the UK’s coastline and southern heathlands.

As the result of anthropogenic developments (e.g., golf course construction and caravan sites), the construction of sea defence structures and the over-stabilisation of sand dune systems (some sand dune systems were reduced to 6% mobility by the 1990’s – Rhind *et al*., 2013), sand lizard populations plummeted to the point that local extinctions were common and the species was lost altogether in Wales in the 1960s (Hill *et al*., 2018).

Since the early 2000s, conservation organisations such as Natural Resources Wales (NRW), Amphibian and Reptile Conservation Trust (ARC), UK zoos and government agencies have successfully been captive-breeding sand lizards for re-introductions in both Wales and England (Moulton *et al*., 2011). As of 2012, 76 official releases using 9000 individual sand lizards had been carried out in England and Wales with a success rate of 65% (defined as a population being successfully established after five years of monitoring) (Blanke & Fearnley, 2015). Re-introductions into both existing populations and areas where local extinctions have occurred have been carried out using sand lizards from the

three ‘races’ of UK sand lizards, from Merseyside, Dorset, and the Weald (ARC Trust, no date).

Using individuals from Merseyside, successful re-introductions have been carried out at multiple locations in North Wales, including Talacre, Morfa Harlech, Aberdyfi/Tywyn and Ynyslas. Moreover, populations have appeared within other sand dune systems where sand lizards have not been recorded previously in Wales, such as Newborough Warren and Aberffraw, both on Anglesey (Hill *et al*., 2018). Given the low probability of an original, native population going undetected in areas such as Newborough Warren, this suggests the occurrence of “unofficial” introductions by amateur herpetologists or herpetoculturists. Unofficial introductions raise concerns such as biosecurity and genetic integrity of sand lizard populations. Given the large distribution of sand lizards across Europe and parts of Asia, there are valid concerns about the origin of such individuals, e.g., herpetoculturists releasing pet lizards with ancestry from outside of the UK.

In 2018, another population was discovered on the Gower Peninsula in South Wales, an area where they were not previously recorded, with up to 20 juvenile lizards being surveyed in under one hour (G. Mee, pers. comm). An early investigation found the specimens in the population to be in generally good condition. It also identified the Gower population as distant from any existing *Lacerta agilis* population and any historical records of the species, and thus outside its presumed ancestral range in Wales by a considerable margin (Edgar, 2007). This, once again, suggests an unofficial, but successful, introduction of a sand lizard population, and raises the question of the origin of this population. If the Gower sand lizards are to be afforded the same conservation importance and legal protection as other established UK populations, it is vital that the origin and ‘race’ of this population is known.

Determination of a non-UK origin for the Gower sand lizards would provoke discussions as to protection status, appropriateness of conservation action and/or means of control/removal.

Purpose of Investigation

Assessing the origin and ancestry of Gower sand lizards is a fundamental step in assessing the need and desirability for conservation action to preserve the population, or, in the case of an introduced population of a non-native lineage, to consider the case for removal. The purpose of this investigation is therefore to use a state-of-the-art whole genome sequencing approach to test the three most plausible hypotheses for the origin of the Gower population of *Lacerta agilis*, namely that it represents:

i. a hitherto overlooked native population

ii. an “unofficial” release of British sand lizard stock by private breeders, or

iii. a colony descended from escaped/released pet lizards from elsewhere in their extensive European and Asian range.

Methods

Sample Acquisition and Sampling Locations

This investigation used 16 buccal swab DNA extractions which had been collected previously as part of a PhD investigation in Bangor University. These samples included those from locations in Wales (Harlech, Gwynedd, and Aberffraw, Anglesey), northwest England (Merseyside), south England (Wareham, Dorset, and Frensham, Surrey) and the Netherlands (Bergherbos). Based on the quality of DNA extracts, four Gower buccal swab samples were selected from a total of 20 donated by Amphibian and Reptile Conservation (ARC). See table 1 for individual sample information. Data for a Danish individual was also available from the European Nucleotide Archive (sample accession: SAMN14333254) at the time of study and was included in our analysis.

Table 1. Individual sample information

Sample ID	Location collected	Sample type	Initial DNA concentration (ng/ul)
wwLA5317	Harlech, Gwynedd	Buccal swab	>600
wwLA5318	Harlech, Gwynedd	Buccal swab	1.17
wwLA5316	Harlech, Gwynedd	Buccal swab	1.59
wwLA567	Harlech, Gwynedd	Buccal swab	5.86
wwLA5319	Aberffraw, Anglesey	Buccal swab	>600
wwLA5320	Aberffraw, Anglesey	Buccal swab	>600
wwLA688	Aberffraw, Anglesey	Buccal swab	5.50
wwLA5340	Gower Peninsula	Buccal swab	5.46
wwLA5341	Gower Peninsula	Buccal swab	4.96
wwLA5339	Gower Peninsula	Buccal swab	4.28
wwLA5336	Gower Peninsula	Buccal swab	2.60
wwLA5322	Merseyside	Buccal swab	5.32
wwLA5323	Merseyside	Buccal swab	7.1
wwLA32	Wareham, Dorset	Buccal swab	2.86
wwLA15	Wareham, Dorset	Buccal swab	1.79
wwLA30	Wareham, Dorset	Buccal swab	1.71
wwLA456	Frensham, Surrey	Buccal swab	4.14
wwLA457	Frensham, Surrey	Buccal swab	3.88
wwLA455	Frensham, Surrey	Buccal swab	2.70
wwLA577	Bergherbos, Netherlands	Buccal swab	6.32
wwLA5349	Potsdam, Germany	Tissue	6.80
DM111	Denmark	published data	N/A

Genomic Sequencing

Method overview

Medium coverage whole genome sequencing involves generating tens of millions of short sequencing reads from a DNA sample of each individual in a random fashion. Sequence reads are then aligned (or computationally "mapped") to the complete reference nuclear genome sequence of the sand lizard, allowing the whole genome of each study individual to be reconstructed and compared. This typically recovers millions of genetic variants encompassing tens of thousands of genetic loci (independently evolving recombining sections of the genome), providing exceptional resolution of population ancestry from only a small number of individuals, as were available to us at the time of study. Mitochondrial sequences can additionally be recovered from the data pool using similar methods, which can be compared with published sequences from phylogeographic studies.

Laboratory work

DNA extraction was carried out using a Qiagen DNeasy Blood and Tissue Kit (https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/dneasy-blood-and-tissue-kit/). Buccal swabs were processed by cutting half of the spongy tip of the swab and treating it as 'tissue' when following the manufacturers protocol. DNA quantification was assessed using a Qubit 3 fluorometer with the dsDNA HS assay kit (https://www.thermofisher.com/order/catalog/product/Q32850#/Q32850), following the manufacturer's protocol.

DNA sequencing by commercial facility

Library preparation, quality control and whole genome sequencing on an Illumina NovaSeq Instrument (2x 150 bp PE reads) to a target 8x coverage was carried out by Novagene Ltd., Cambridge (https://en.novogene.com/services/research-services/genome-sequencing/whole-genome-sequencing/?gclid=CjwKCAiA1JGRBhBSEiwAxXblwef3iO7lvfRubQkFxSrhAMjAg_-nQlkoI7XxFKqG2Y3m6HZwoZVxNhoCRNkQAvD_BwE#). Four samples (two from Harlech, one from Aberffraw and one from Frensham) were excluded after Novagene's initial quality control due to high levels of DNA fragmentation and one sample was excluded after library preparation quality control (Wareham) resulting in 15 whole genome datasets to 8x coverage. See table 2.

Table 2. Individual sample status following initial and library preparation quality control by Novagene Ltd.

Sample ID	Initial quality control	Library preparation quality control
wwLA5317	Passed	Passed
wwLA5318	Passed	Passed
wwLA5316	Failed	

wwLA567	Failed	
wwLA5319	Passed	Passed
wwLA5320	Passed	Passed
wwLA688	Failed	
wwLA5340	Passed	Passed
wwLA5341	Passed	Passed
wwLA5339	Passed	Passed
wwLA5336	Passed	Passed
wwLA5322	Passed	Passed
wwLA5323	Passed	Passed
wwLA32	Passed	Passed
wwLA15	Passed	Passed
wwLA30	Passed	Failed
wwLA456	Passed	Passed
wwLA457	Passed	Passed
wwLA455	Failed	
wwLA577	Passed	Passed
wwLA5349	Passed	Failed
DM111	N/A	N/A

Data Processing

Initial data processing involved trimming adapter sequences (synthetic DNA sequences that can occur as part of the sequencing process) from PE reads, requiring a single base overlap between adapter and read (setting "-O 1"), and removing short reads < 30 bp in length (setting -m 30), using the program Cutadapt v. 1.18 (Martin 2011). Overlapping PE read pairs were then identified and merged using the program FLASH v. 1.2.11 (Magoč & Salzberg 2011) using default parameter settings. The merged and unmerged PE reads were then mapped to the reference genome assembly of the sand lizard (rLacAgil1: https://www.ncbi.nlm.nih.gov/assembly/GCF_009819535.1/) in separate analyses using the "mem" algorithm of the program bwa v. 0.7.17(Li & Durbin 2009), using default parameter settings. The program samtools v. 1.3.1 (Li *et al*. 2009) was then used to filter out unmapped reads (setting "-F 4"), secondary mapped reads (setting "-F 256") and reads with low mapping quality (setting "-q 30"). Samtools was then used to sort reads by their 5' mapping position (command "sort") and remove potential PCR duplicates (command "rmdup"). Finally, samtools was used to combine the merged and unmerged PE read data into a single file in bam format (command "merge") for further analysis.

PCA and genetic distance analysis

ANGSD version 0.925 (Korneliussen et al. 2014) was used to calculate distance and covariance matrices for the analysis of relatedness between our 16 genome datasets using principal component analysis (PCA) and neighbour-joining phylogenetic analysis (NJ). Further detailed comparisons of pairwise genetic distances were also made from the genetic distance matrix.

The distance and covariance matrices were generated simultaneously by calling a majority-rule consensus base for each position of the reference genome, for each individual (setting “doIBS 2”). The following filters were applied: base quality (setting “-minQ 30”), mapping quality (setting “-minMapQ 30”), minimum individual read depth (setting “-setMinDepthInd 4”), require data from >1/3 of individuals (setting “-minInd 6”), restrict analysis to only chromosomal scaffolds (setting “-rf”). PCA was carried out using the “eigen” function in R (R Core Team 2014). The neighbour joining phylogeny was calculated using the “ape” library (Paridis et al. 2004) in R, and rooted using the Danish individual.

Mitochondrial sequence recovery and analysis

Data were additionally mapped to the mitochondrial genome sequence included as part of the rLacAgi1 genome assembly, using identical methods as described above except that the bwa “aln” algorithm was utilised, and reads were only filtered for low mapping quality. A consensus sequence for the cytochrome b gene region was then generated using angsd (setting “-doFasta 2” specifying gene coordinates using setting “-r”), with leading and trailing Ns trimmed using the program SeqKit (Shen *et al*., 2016). These cytb sequences were analysed using MEGA (version 11.0.10 – Tamura *et al*., 2021) and aligned (using default MEGA settings) alongside 52 sequences from sand lizard subspecies found throughout their mainland distribution (Andres *et al*., 2014) and a Bulgarian green lizard (*Lacerta viridis*) outgroup (Böhme *et al*., 2007). A neighbour joining tree (maximum composite likelihood, transition + transversion substitutions included, uniform rates among sites, homogenous patterns among lineages and pairwise deletion) was constructed using MEGA. In an attempt to dissect within-UK relationships, cytb minimum network haplotype analysis was carried out using Popart (version QT 4.8.5 – Leigh & Bryant, 2015).

Heterozygosity and runs of homozygosity

Genetic health was assessed using two measures. Average heterozygosity was used to provide a measure of overall genetic diversity. We also carried out an inspection for large genomic runs of homozygosity, which can result when an individual’s parents are close relatives and is therefore indicative of inbreeding (Curik et al., 2014).

We restricted our analyses to chromosome 1 due to the time constraints of the project. Chromosome 1 is an autosome spanning approximately 130 megabases. We calculated the site frequency spectrum for each individual in angsd relative to the reference genome assembly (command “-doSaf”). Base and mapping quality filters we used as described previously, and we limited our analysis to sites between half and 2x the average genome wide coverage (typically between 5 and 20 reads). A maximum likelihood estimation was then carried out using the realSFS program (distributed with angsd) along a 1 megabase non-overlapping sliding window for chromosome 1. Heterozygosity was calculated by dividing the number of biallelic sites by the total number of sampled sites, and the results plotted in R.

Results

Data production

Of the 16 WGS mapped to the sand lizard genome, one sample (wwLA15) was excluded from further analysis due to low mapping coverage (1.35609 Gb). The Danish sequence (DM111) had the lowest mapping coverage of 4.76924 Gb as expected and all other sequences had a mapping coverage above 11 Gb. See table 3.

Table 3. Number of mapped bases (Gb) for each individual sample

Sample ID	Mapped bases (Gb)
wwLA5317	12.65796
wwLA5318	14.29139
wwLA5319	11.41338
wwLA5320	13.63293
wwLA5340	15.73291
wwLA5341	14.42245
wwLA5339	14.14035
wwLA5336	12.49645
wwLA5322	12.55360
wwLA5323	12.78358
wwLA32	13.32374
wwLA15	1.35609
wwLA456	12.89592
wwLA457	13.97750
wwLA577	11.92781
DM111	4.76924

Mitochondrial sequence analysis

Phylogenetic analysis of mitochondrial cytb genes demonstrated that haplotypes from the UK, including Gower, and the Netherlands form a clade that is sister to a clade of haplotypes from Central and Eastern mainland Europe (Denmark, Slovakia, Germany, Hungary) (figure 1). Minimum network haplotype network analysis (figure 2) of these two clades showed that all sampled UK sand lizards, including those from Gower, have the same unique cytb haplotype. On this basis we can conclude that the Gower population in all likelihood has a UK matrilineal ancestry. However, mitochondrial DNA does not provide sufficient resolution to differentiate among UK populations, and cannot in itself test for the presence of admixture between lineages, highlighting the requirement for nuclear DNA analysis.

Figure 1. Neighbour-joining tree constructed from mitochondrial cytochrome b sequences produced in this investigation alongside an additional 52 cytochrome b sequences from sand lizard subspecies found in their European mainland range (Andres *et al.*, 2014). A UK clade is highlighted within the red box and Gower sand lizard sequences are contained within the green boxes demonstrating a UK matrilineal ancestry.

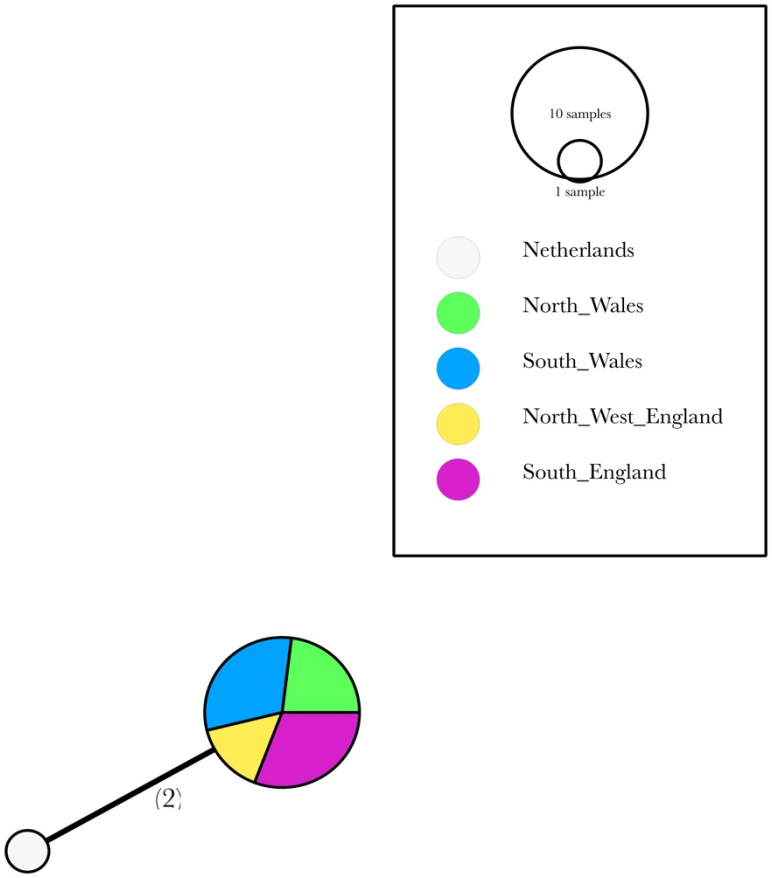

Figure 2. Minimum network haplotype network analysis of cytochrome b sequences of UK sand lizards and one sequence from the Netherlands. All UK sand lizards share one single exclusive cytochrome b haplotype. This indicates the likelihood of UK matrilineal ancestry of the Gower Peninsula and highlights the need for nuclear genetic analysis to further investigate genetic relationships between UK sand lizard populations.

Principal components analysis

PCA of the whole genome datasets which contained a total of 395,413,986 single nucleotide polymorphisms (SNPs) was successfully able to differentiate population clusters within UK sand lizards, and between UK and Danish and Netherlands sand lizards (figure 3). Ordination of individuals along PC1 (16.6% sampled variation) and PC2 (12.5% sampled variation) reveals a tight cluster of individuals, comprising of the Merseyside, Harlech, Aberffraw and Gower sand lizards. The two individuals from Frensham cluster closely together, with the Wareham individual showing a close affinity to this cluster. Mainland European sand lizards are distinct from the UK populations, with the Netherlands sand lizard showing a somewhat closer affinity to the southern UK populations.

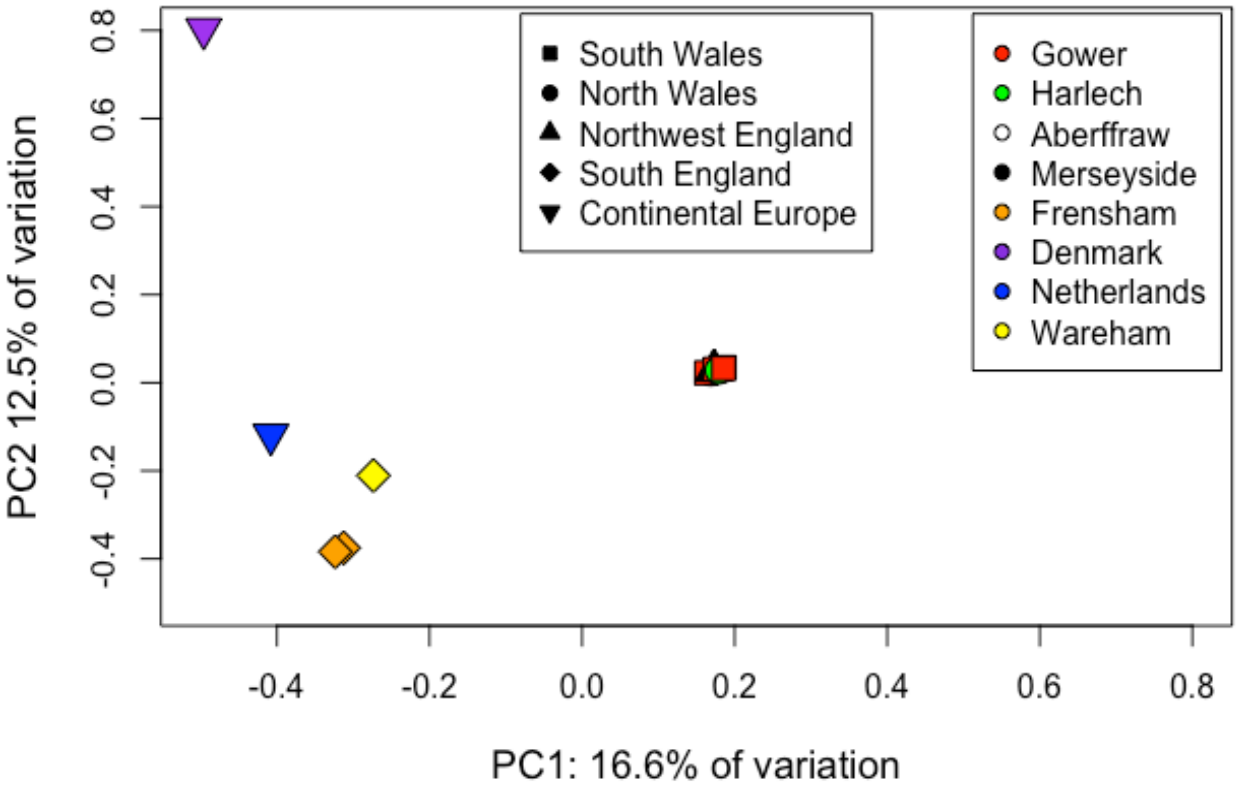

Figure 3. Principal component analysis using the whole genome datasets (UK, Danish and Netherlands) of sand lizards in this investigation. When plotted against PC1 (16.6% of sampled variation) and PC2 (12.5% of sampled variation), a cluster comprising of north and south Wales (Gower) and northwest England (Merseyside) is revealed. Frensham forms another cluster with Wareham showing a close affinity and a distinction to continental Europe is also revealed.

The clustering of Merseyside and Harlech populations is consistent with their known history of source/donor site for the captive breeding and release program. The two Welsh populations resulting from “unofficial” releases, Aberffraw and Gower, share much more recent genetic ancestry with these two north-western populations than with any other sampled UK or mainland European population.

Whole genome phylogenetic analysis

Whole genome phylogenetic analysis was able to differentiate UK sand lizards as a separate clade from continental sand lizards (figure 4). Within the UK clade, we further identify the southern individuals from Frensham and Wareham as a sister to a well-defined clade comprising all sand lizards from Merseyside, North Wales, and the Gower.

Overall, this result is consistent with the PCA in identifying Merseyside, North Wales and Gower sand lizards in this investigation as forming a distinctive, yet homogeneous cluster of animals likely originating from Merseyside ‘dune race’ ancestry.

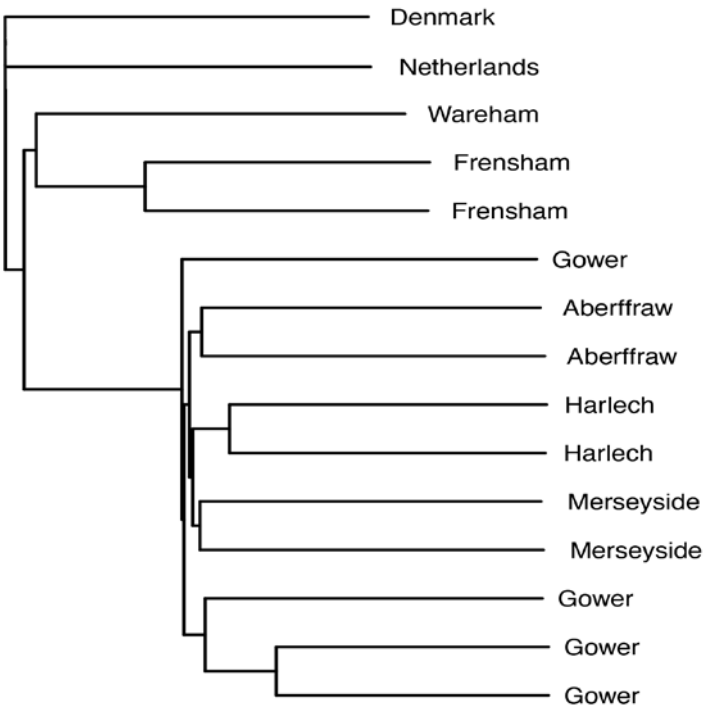

Figure 4. Neighbour-joining whole genome phylogeny of UK, Denmark and Netherlands sand lizards. A north Wales, Gower and Merseyside clade is sister group to a south England clade and both continental sand lizards sit outside a larger UK sand lizard clade.

We further note that within the Merseyside/Wales clade, individuals from the same location cluster together, except in the case of the Gower, where one individual is a phylogenetic outlier. This apparent anomaly may represent a sampling artefact: four Gower individuals were sampled in comparison to two for all other locations, reducing the potential to detect rarer genetic variants in the latter. The branch lengths separating all individuals of the Merseyside/Wales clade are similar to those separating the two Frensham samples, supporting the likely recent origin of the entire clade from a single ancestral stock from Merseyside

Analysis of pairwise genetic distances

More detailed analysis of pairwise genetic distances reveals several important additional facets of the data (figure 5). The key result is that levels of pairwise divergence between individuals of different dune population are no greater than many pairwise distances among individuals *within* dune populations, indicating that the dune populations of the Merseyside/Wales clade are, genetically speaking, best regarded as a single entity.

In terms of genetic distances between individuals from different populations, we found that distances between Gower sand lizards and those from other dune populations and the

south England populations fall within the range observed for the other dune populations. Thus, in terms of overall patterns of genetic divergence, the Gower population appears to be “behaving” as a dune race population, and not as a descendent from a more distant population which is obscured from other analyses as an effect of under sampling.

A result of note is that genetic distances between southern UK and mainland European individuals are consistently lower than those observed between dune race and mainland European individuals. This is consistent with the phylogeny, where a long branch separates the dune race from southern English and mainland populations. This long branch may reflect a founder effect but this would require further investigation. It does, in any case, emphasise the distinctive genetic status of the north-western dune race.

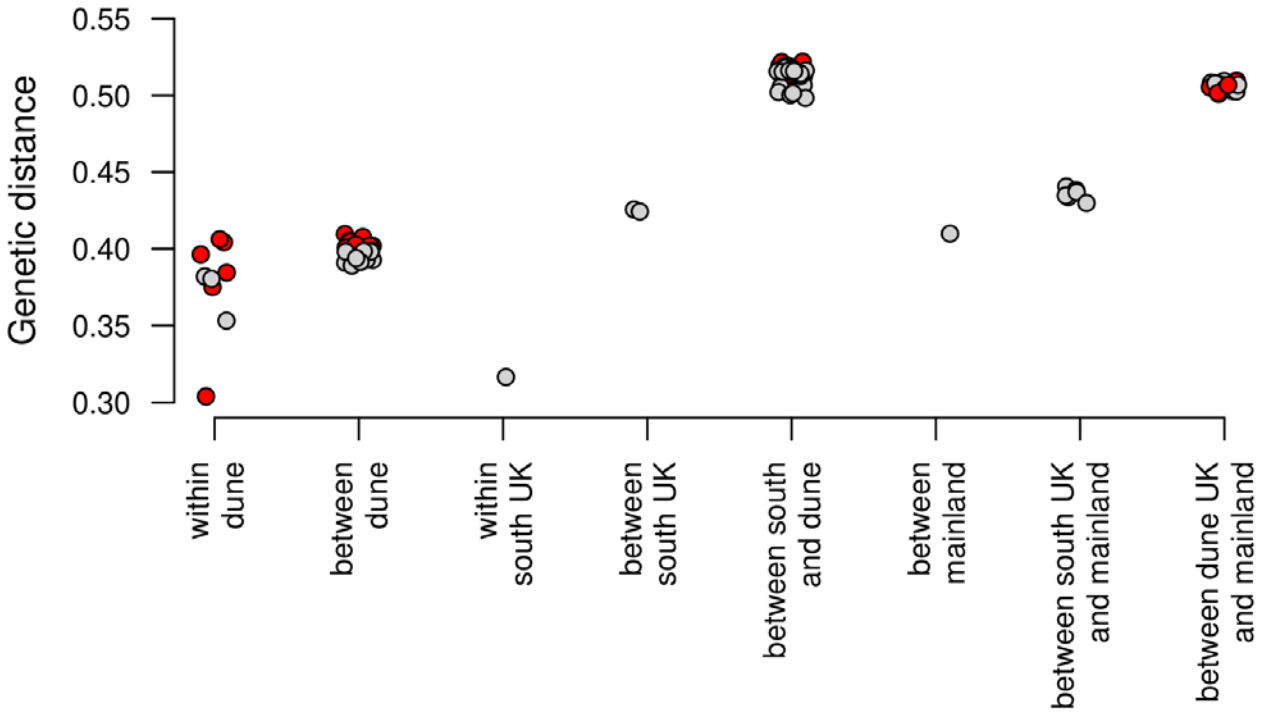

Figure 5. Pairwise genetic distances (y axis) between sand lizard genomes, arranged into pertinent categories discussed in the text (x axis). Each individual point corresponds to a different pairwise comparison involving different individuals, and all possible combinations of individuals within each category are displayed. Comparisons involving Gower individuals are shown in red, and all other comparisons in grey. Note the absolute genetic distance values are the fraction of variable sites screened from the total genome datasets, disregarding invariant positions, and are therefore much higher than the actual genome divergence will be.

Heterozygosity and runs of homozygosity

The estimated proportions of heterozygous sites among the investigated individuals ranged from 0.0057 to 0.0076 (Figure 6). Overall, these values are remarkably high. European humans, for example, show values around ten times lower (0.0006) and chimpanzee heterozygosity, a species generally considered to be genetically diverse, is

typically around 0.001 (Westbury *et al*. 2018). Unfortunately, comparative data from other squamate reptiles is currently lacking, preventing a more taxonomically focussed comparison.

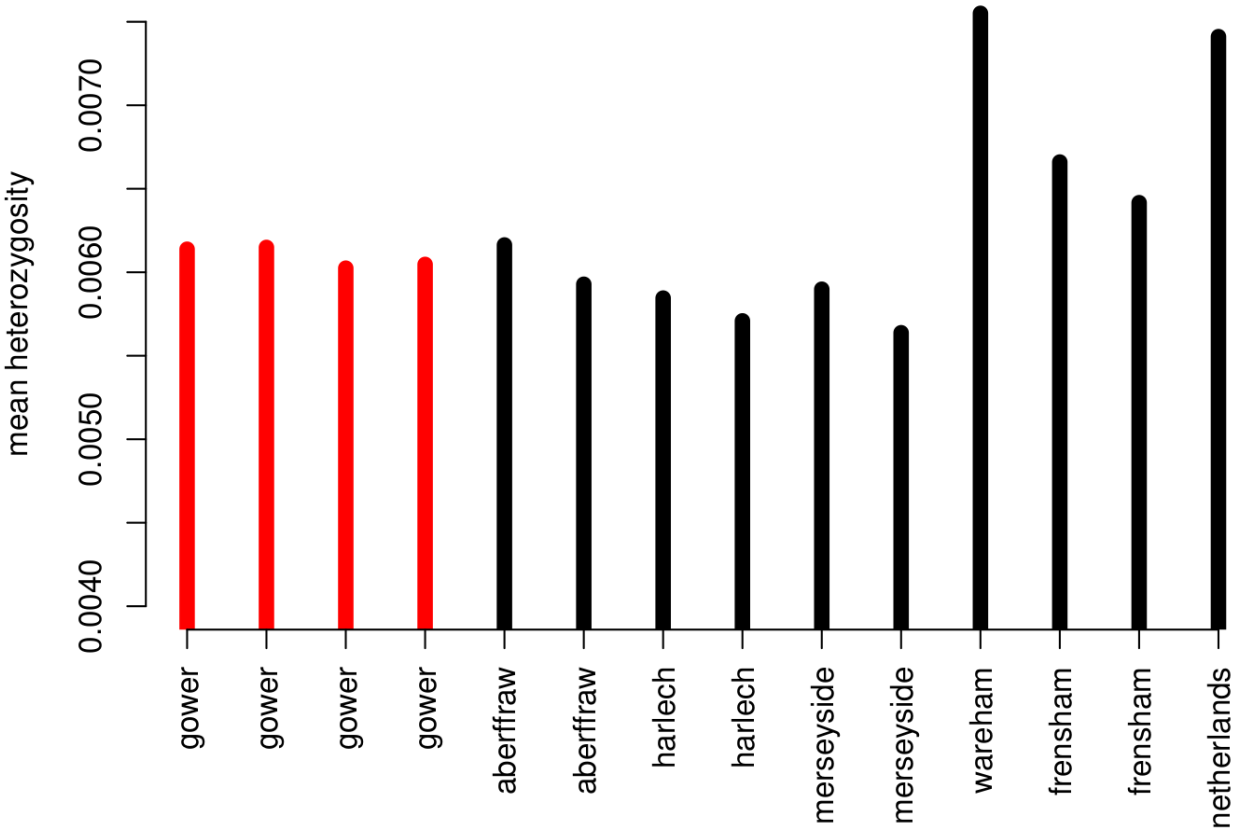

Figure 6. Mean proportion of heterozygous sites (y axis) for the investigated sand lizards (x axis). Values are based on analysis of Chromosome 1 (~130 megabases). Results for the Gower individuals are shown in red.

We do observe that the mean heterozygosity for the Gower individuals is approximately equal to, or slightly higher than, that observed in other dune race animals. Interestingly the lowest observed heterozygosity occurs in a Merseyside animal from the presumed source population from which all others were founded, potentially indicating an ongoing decline in diversity in Merseyside since the establishment of the captive breeding programme. Southern England populations have consistently higher heterozygosity than all dune race individuals, being highest in the Wareham population and intermediate in Frensham. The Netherlands lizard's heterozygosity is also high, and close to that of the Wareham individual.

Inspection of heterozygosity values along sliding windows revealed several instances of runs of homozygosity in both introduced (figure 7) and natural (figure 8) populations. We applied a post-hoc definition of a contiguous 3 megabase section of extremely and uniformly low heterozygosity as an appropriate diagnostic criterion. This suggested the presence of runs of homozygosity in one Aberffraw individual, one Harlech individual, one

Merseyside individual, the Netherlands individual, and both Frensham individuals. All these populations therefore appear to be suffering the effects of inbreeding. Frensham in particular shows quite extreme effects, with one individual showing three runs of homozygosity, and the second individual showing the largest observed run of homozygosity across all individuals, spanning 10 megabases and representing around 7.5% of the entire chromosome 1. Notably no evidence of runs of homozygosity was observed in any Gower individual. Their genetic health can therefore be considered high in comparison to UK sand lizards as a whole.

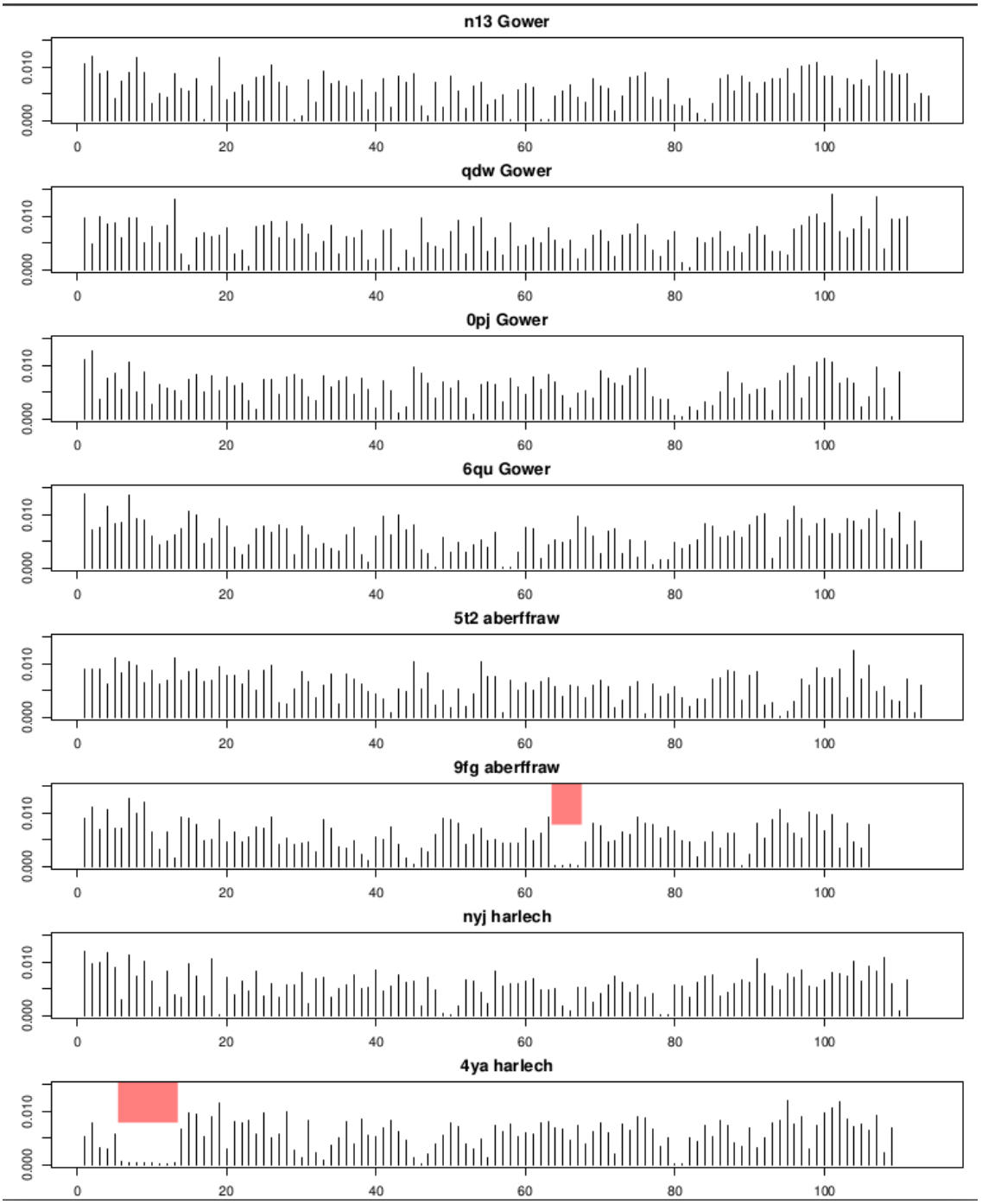

Figure 7. Runs of homozygosity in introduced sand lizard populations. The x axis represents chromosome 1 with the position indicated in megabases (of around 130 megabases total length). Individual bars show the proportion of heterozygous positions for a 1 megabase window (Y axes, all scaled between 0–0.015). Runs of homozygosity are highlighted in red. The length of the sampled chromosome differs slightly between individuals due to variable coverage and the effects of filtering, but this represents a maximum of only 8 windows between individuals and can be generally ignored.

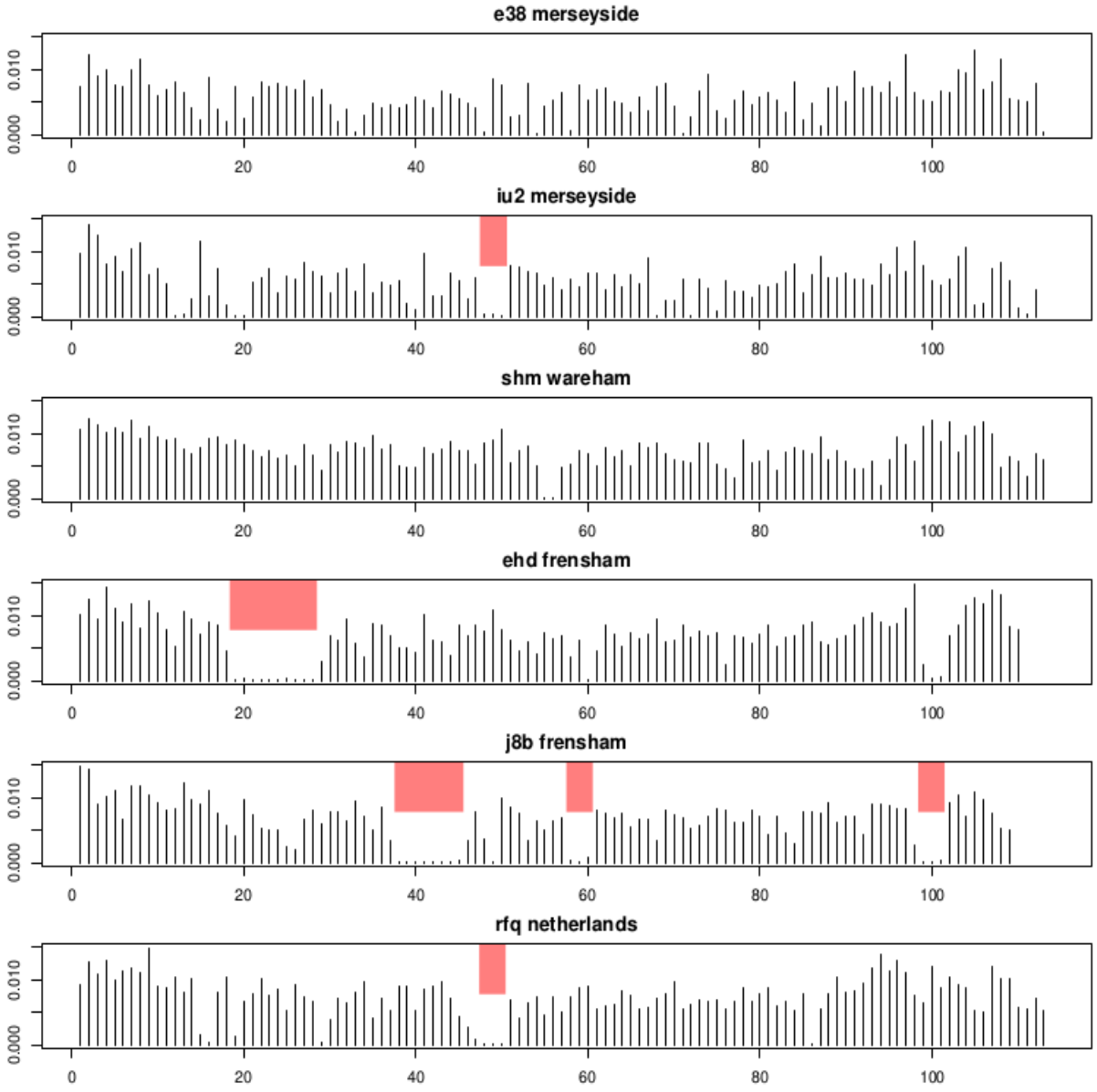

Figure 8. Runs of homozygosity in natural sand lizard populations. For details see Figure 7 legend.

Discussion

Ex-situ breeding and reintroductions of species to suitable habitats can be a successful approach to improving the conservation status of a species. The sand lizard (*Lacerta agilis*) in the UK is a striking example of a successful ex-situ conservation initiative, with reintroduced populations now making up a significant proportion of the known populations, including its entire Welsh range (Blanke & Fearnley, 2015).

However, a key concern in reintroduction programmes is ensuring the pedigree of the founder stock. This is essential to ensure the genetic health of the reintroduced population by avoiding inbreeding depression, but also to avoid disrupting locally adapted gene constellations that confer selective advantages in the local habitat, and the homogenisation of the global population (Witzenberger & Hochkirch, 2011; Champagnon *et al*., 2012). Consequently, constant genetic monitoring of the captive population destined for reintroductions is required.

In the case of *Lacerta agilis* in the UK, "new" populations not established through officially sanctioned release programmes have been discovered in several locations such as Aberffraw and Newborough Warren on Anglesey (Hill *et al*., 2018), and, most recently, on the Gower Peninsula (Donald et al., 2021; G. Mee, pers. comm.). The Aberffraw population appears to be the result of an unsanctioned release of Merseyside stock (Russell, 2012), and the Newborough Warren population is similarly likely to be the result of an "unofficial" introduction, but the origin of the geographically distant Gower population had been entirely unresolved until now.

This phenomenon of unofficial introductions raises questions on the genetic make-up of the introduced populations: if these originated from mainland Europe, there would be a risk of compromising the integrity of the gene pool of native UK sand lizards (as has happened in other UK herpetofauna – e.g., Brede, 2015), and releasing admixed individuals, even of UK origin, could potentially disrupt adaptive processes in different habitats. Consequently, understanding the pedigree of unexpected populations of sand lizard discovered in the UK is essential to inform management decisions on the fate of these populations.

Genetic approaches play a key role in identifying the ancestry of sand lizard populations. In the recent past, microsatellites were the method of choice in conservation genetics. However, these require very large sample sizes and offer limited resolution from the small number of loci involved (typically around 10-20 – Russell, 2012). In recent years, Whole Genome Sequencing (WGS) methods have become affordable for conservation genetic work. Due to the very large number of loci analysed, the methods can provide very high-resolution population data based on small sample sizes and low-quality DNA (Węcek *et al*., 2017).

The use of WGS approaches in this study has allowed us to achieve the following:

1. We have been able to confirm that the focal population, the recently discovered sand lizards from the Gower Peninsula, are of UK ancestry and indistinguishable from north-western dune lizards from Merseyside or descended introduced populations in North Wales.

2. We have confirmed the "dune race" of sand lizard from north-western England (Merseyside), and populations in North Wales known to be descended from them (Morfa Harlech), as a genetically distinct management unit within the UK sand lizard clade.

3. We have confirmed the results of Russell (2012), who found the "unofficial" Aberffraw population to be descended from Merseyside stock

4. The lack of differentiation of the Gower population from other dune sand lizards indicates that it was most likely introduced "unofficially" in the shape of stock

descended from Merseyside sand lizards or populations derived from Merseyside stock.

5. We have confirmed that the Gower population is in good genetic health, with high levels of heterozygosity and no indication of runs of homozygosity that would be indicators of inbreeding. This contrasts to the situation in several other UK populations, where there is evidence of inbreeding.

1) 2)

Plate 1. Male sand lizards. 1) Specimen from Sefton Coast, Merseyside. 2) Specimen from the Gower population. Note the typical pattern of the dune race, with largely uniform green sides in males, common to both populations © Pete Hill

From a conservation perspective, our data provide additional grounds for treating the north-western dune race of the UK sand lizard as a separate management unit from southern heathland *Lacerta agilis*. The “unofficial” introductions at Aberffraw and on the Gower consist of apparently pure dune race stock. As such, while these may have been introduced in places without historical records of sand lizards, they are genetically indistinguishable from the native population on Merseyside and the Morfa Harlech population, reintroduced to North Wales based on captive-bred Merseyside stock. These unofficially introduced populations appear to be genetically healthy and thus contribute positively towards the favourable conservation status of the sand lizard in the UK in general and in Wales in particular, and specifically to that of the highly threatened dune race. The genetic identity of the Gower sand lizards as a pure dune race population should inform conservation policy towards this recently discovered population. In particular, it should be taken into account for European Protected Species licencing and targeted management prescriptions in the area of occurrence.

Conclusion & Considerations

This investigation has not only convincingly answered the question of the origin and genetic health of the Gower sand lizards, but it has also confirmed the usefulness of WGS

techniques to answer questions on the origin and pedigree of populations of uncertain ancestry at an affordable cost, with very limited sampling effort and with high precision.

The data generated by this approach are entirely transferrable and unbiased (unlike microsatellite markers). The genomic resources generated as part of this study thus provide a robust basis for similar investigations of other unexpected occurrences of sand lizards in the UK.

References

Andres C, Franke F, Bleidorn C, Bernhard D, & Schlegel M. 2014. Phylogenetic analysis of the *Lacerta agilis* subspecies complex. Systematics and Biodiversity 12 (1), 43-54.

Amphibian and Reptile Conservation (ARC). No date. Saving species: Sand lizards. Available from: https://www.arc-trust.org/saving-sand-lizard [accessed on 01/03/2022]

Blanke I & Fearnley H. 2015. The Sand Lizard: Between Light and Shadow. Bielefeld: Laurenti Verlag.

Böhme MU, Fritz U, Kotenko T, Džukić G, Ljubisavljević K, Tzankov N, Berendonk TU. 2007. Phylogeography and cryptic variation within the *Lacerta viridis* complex (Lacertidae, Reptilia). Zoologica Scripta 36, 119–131.

Brede E. 2015. Beam Brook revisited: a molecular study of a historically introduced non-native amphibian (*Triturus carnifex*) and its limited introgression into native UK *Triturus cristatus* populations. Amphibia-Reptilia 36, 287–299.

Champagnon J, Elmberg J, Guillemain M, Gauthier-Clerc M., Lebreton J-D. 2012. Conspecifics can be aliens too: A review of effects of restocking practices in vertebrates. Journal for Nature Conservation 20, 231–241.

Curik I, Ferenčaković M., Sölkner J. 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science* 166, 26–34.

Donald HCF, Foster J, Wilkinson JW, Hill P, Barber M, Mee G, Edgar P, Marschang RE, Sainsbury AW. 2021. Two novel adenoviruses in free-living British lizards. EcoHealth 18, 297–300.

Edgar P. 2007. The Conservation Status of the Natterjack Toad *Bufo calamita* and Sand Lizard *Lacerta agilis* in Wales. Contract Science Report No: 788. CCW, Bangor.

Hill P, Moulton N & Foster J. 2018. Sand lizard surveys at Newborough Warren NNR and sand dune habitat management guidance. NRW Evidence Report No: 302, 103 pp, Natural Resources Wales, Bangor.

Korneliussen, TS, Albrechtsen A, & Nielsen R. 2014. "ANGSD: analysis of next generation sequencing data." BMC bioinformatics 15.1, 1-13.

Leigh JW, Bryant D. 2015. PopART: Full-feature software for haplotype network construction. Methods Ecol Evol 6 (9), 1110–1116.

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis, G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16), 2078–2079.

Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics (Oxford, England), 27(21), 2957–2963.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. Journal *17*(1), 10.

Moulton N, Wilkinson J, Davis C, Foster J & Howe L. 2011. Sand lizard translocation in the UK. Global Re-introduction Perspectives: 2011. More case studies from around the globe, Re-introduction Specialist Group and Abu Dhabi, UAE: Environment Agency-Abu Dhabi, 116-119.

R Core Team. 2014. R: A language and environment for statistical computing (R Found. Stat. Comput.), Available at. http://www.r-project. org/

Rhind P, Jones R & Jones L. 2013. The Impact of Dune Stabilization on the Conservation Status of Sand Dune Systems in Wales. Restoration of Coastal Dunes. Springer Series on Environmental Management. Springer, Berlin, Heidelberg, 125-143.

Russell L. 2012. The Conservation and Landscape Genetics of the Sand Lizard *Lacerta agilis*. Unpublished PhD thesis, University of Sussex.

Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS ONE, 11(10), e0163962.

Soorae PS. 2011. Global Re-introduction Perspectives: 2011. More case studies from around the globe, Gland, Switzerland: IUCN/SSC Re-introduction Specialist Group and Abu Dhabi, UAE: Environment Agency-Abu Dhabi.

Tamura K, Stecher G, and Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis version 11. Molecular Biology and Evolution 38, 3022–3027.

Węcek K, Hartmann S, Paijmans JLA, Taron U, Xenikoudakis G, Cahill JA, Heintzman PD, Shapiro B, Baryshnikov G, Bunevich AN, Crees JJ, Dobosz R, Manaserian N, Okarma H, Tokarska M, Turvey ST, Wójcik JM, Zyla W, Szymura JM, Hofreiter M, Barlow A. 2016. Complex Admixture Preceded and Followed the Extinction of Wisent in the Wild. Molecular Biology and Evolution, 34(3), 598-612.

Westbury MV, Hartmann S, Barlow A, Wiesel I, Leo V, Welch R, Parker DM, Sicks F, Ludwig A, Dalén L, Hofreiter M. 2018. Extended and continuous decline in effective

population size results in low genomic diversity but no signs of inbreeding in the world's rarest hyena species, the brown hyena. Molecular Biology and Evolution 35, 1225–1237.

Witzenberger KA, Hochkirch A. 2011. Ex situ conservation genetics: a review of molecular studies on the genetic consequences of captive breeding programmes for endangered animal species. Biodiversity and Conservation 20, 1843–1861.

Appendices

Data Archive Appendix

At the time of submission all genetic data files are stored on the Supercomputing Wales system "Hawk", physically located at Cardiff University, with additional local back-up copies stored in the Environment Centre Wales, Bangor University. The data are in fastq file format and are compressed using gzip compression. The total file size is ~165 Gb, noting the uncompressed file size will be much larger. The files cannot be opened using standard Windows-based computing systems, and doing so may result in data corruption.

As a short term solution, any individuals requiring access to the data can contact the authors who can provide copies or access to Supercomputing Wales based on individual needs and circumstances. In the longer term these data will be uploaded to a public database (e.g. European Nucleotide Archive) following standard practise in the genetics field. An overview of the raw data files is shown in Table X.

Table 4. Raw data files generated during this project.

Sample	File name	Size
wwLA15	peg+s15_EDSW210026924-1a_HV7J3DSX2_L2_R1.fastq.gz	5.5G
wwLA15	peg+s15_EDSW210026924-1a_HV7J3DSX2_L2_R2.fastq.gz	5.7G
wwLA32	shm+s32_EDSW210026923-1a_HV7CFDSX2_L3_R1.fastq.gz	5.4G
wwLA32	shm+s32_EDSW210026923-1a_HV7CFDSX2_L3_R2.fastq.gz	5.7G
wwLA456	ehd+s456_EDSW210026926-1a_HV7J3DSX2_L3_R1.fastq.gz	5.0G
wwLA456	ehd+s456_EDSW210026926-1a_HV7J3DSX2_L3_R2.fastq.gz	5.2G
wwLA457	j8b+s457_EDSW210026927-1a_HV73WDSX2_L2_R1.fastq.gz	5.5G
wwLA457	j8b+s457_EDSW210026927-1a_HV73WDSX2_L2_R2.fastq.gz	5.8G
wwLA5317	nyj+s5317_EDSW210026920-1a_HV7CFDSX2_L3_R1.fastq.gz	4.8G
wwLA5317	nyj+s5317_EDSW210026920-1a_HV7CFDSX2_L3_R2.fastq.gz	5.0G
wwLA5318	4ya+s5318_EDSW210026922-1a_HV7J3DSX2_L2_R1.fastq.gz	5.7G
wwLA5318	4ya+s5318_EDSW210026922-1a_HV7J3DSX2_L2_R2.fastq.gz	6.0G
wwLA5319	5t2+s5319_EDSW210026929-1a_HTVWGDSX2_L2_R1.fastq.gz	3.6G
wwLA5319	5t2+s5319_EDSW210026929-1a_HTVWGDSX2_L2_R2.fastq.gz	3.7G
wwLA5319	rb0+s5319_EDSW210026929-1a_HV5LKDSX2_L4_R1.fastq.gz	459M

wwLA5319	rb0+s5319_EDSW210026929-1a_HV5LKDSX2_L4_R2.fastq.gz	482M
wwLA5320	9fg+s5320_EDSW210026930-1a_HTVWGDSX2_L1_R1.fastq.gz	5.1G
wwLA5320	9fg+s5320_EDSW210026930-1a_HTVWGDSX2_L1_R2.fastq.gz	5.3G
wwLA5322	iu2+s5322_EDSW210026918-1a_HV7J3DSX2_L3_R1.fastq.gz	4.6G
wwLA5322	iu2+s5322_EDSW210026918-1a_HV7J3DSX2_L3_R2.fastq.gz	4.9G
wwLA5323	e38+s5323_EDSW210026919-1a_HV7C2DSX2_L1_R1.fastq.gz	4.7G
wwLA5323	e38+s5323_EDSW210026919-1a_HV7C2DSX2_L1_R2.fastq.gz	4.9G
wwLA5336	6qu+s5336_EDSW210026916-1a_HV7C2DSX2_L1_R1.fastq.gz	4.7G
wwLA5336	6qu+s5336_EDSW210026916-1a_HV7C2DSX2_L1_R2.fastq.gz	4.8G
wwLA5339	qdw+s5339_EDSW210026915-1a_HV7C2DSX2_L1_R1.fastq.gz	5.1G
wwLA5339	qdw+s5339_EDSW210026915-1a_HV7C2DSX2_L1_R2.fastq.gz	5.3G
wwLA5340	n13+s5340_EDSW210026913-1a_HV7C2DSX2_L4_R1.fastq.gz	5.9G
wwLA5340	n13+s5340_EDSW210026913-1a_HV7C2DSX2_L4_R2.fastq.gz	6.1G
wwLA5341	0pj+s5341_EDSW210026914-1a_HV7C2DSX2_L1_R1.fastq.gz	5.3G
wwLA5341	0pj+s5341_EDSW210026914-1a_HV7C2DSX2_L1_R2.fastq.gz	5.5G
wwLA577	r4q+s577_EDSW210026932-1a_HV5TVDSX2_L2_R1.fastq.gz	4.6G
wwLA577	r4q+s577_EDSW210026932-1a_HV5TVDSX2_L2_R2.fastq.gz	4.8G

Data outputs associated with this project are archived in the Supercomputing Wales system “Hawk”, physically located at Cardiff University, with additional local back-up copies stored in the Environment Centre Wales, Bangor University on server–based storage at Natural Resources Wales.

The data archive contains:

[A] The final report in Microsoft Word and Adobe PDF formats.

[B] A series of Raw Data files in a Fastq format.

Metadata for this project is publicly accessible through Natural Resources Wales’ Library Catalogue https://libcat.naturalresources.wales (English Version) and https://catllyfr.cyfoethnaturiol.cymru (Welsh Version) by searching ‘Dataset Titles’. The metadata is held as record no 125331.

Further copies of this report are available from library@cyfoethnaturiolcymru.gov.uk

image: Ellie Chidlow

For more information, visit **www.dynamicdunescapes.co.uk** or contact the Dynamic Dunescapes team on **dynamicdunescapes@naturalengland.org.uk**