SUPPLEMENTARY MATERIALS

Selection on the vascular-remodelling *BMPER* gene is associated with altitudinal adaptation in an insular lizard

Nina Serén[1,2,3], Catarina Pinho[1,2], Rodrigo Megía-Palma[1,2,4], Prem Aguilar[1,2,3], Anamarija Žagar[1,2,5,6], Pedro Andrade[1,2], Miguel A. Carretero[1,2,3]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Geneticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal
[2]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
[3]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4099-002 Porto, Portugal
[4]Universidad de Alcalá (UAH), Parasitology Unit, Department of Biomedicine and Biotechnology, School of Pharmacy, Spain
[5]Department of Organisms and Ecosystems Research, National Institute of Biology, Večna pot 111, 1000 Ljubljana, Slovenia
[6]Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia

Corresponding author: Nina Serén (ninagseren@cibio.up.pt)

Supplementary text – Methods

(a) Field sampling

Samples of *Gallotia galloti* were collected in two field missions held between July and August of 2017 and 2018, in different locations in the island of Tenerife, including high and low-altitude sites along two main transects (Table S1). An additional individual from Parque Nacional del Teide (28.255,-16.621; 2,300 m elevation) was collected for reference genome assembly. Lizards were caught either by pitfall traps using tomato and banana as bait or by noose. Only males were sampled to maximise coverage in scaffolds derived from the Z chromosome. Biological samples consisted of tail tip tissue, minimising stress and disturbance (García-Muñoz et al., 2011; Langkilde & Shine, 2006) and taking advantage of the lizards' fast regeneration capacity (Fernández-Rodríguez et al, 2021). Moreover, < 50µL of blood was extracted from the caudal tail vein with sterile needles (BD microlance 3; 23G; 0.6 × 25 mm) from a previously ethanol-disinfected skin area and more than 2 cm away from the cloaca to avoid damaging the hemipenises. Samples were immediately stored in 96% ethanol until DNA extraction. Lizards were released in the same location of trapping after ecophysiological testing.

(b) Reference genome assembly

To assemble the genome of *G. galloti de novo* we used Chromium linked read technology from 10X Genomics (Weisenfeld et al., 2017). Muscle tissue from a single male was sent to the Genomics Services Laboratory of the HudsonAlpha Institute for Biotechnology, where genomic DNA was extracted using the MagAttract HMW DNA Kit (QIAGEN). Following fragment size evaluation, a Chromium library was prepared following the manufacturer's protocol, and sequenced on a NovaSeq instrument (Illumina). This generated a total of 600.03 million reads. To assemble the genome, we used *Supernova* v2.1.1 (Weisenfeld et al., 2017) with the full read dataset (raw coverage of 69.28X) and default options. An additional assembly was trialled with a reduced number of reads (raw coverage ~35X) but this resulted in an assembly with a large number of duplicate scaffolds, so this was discarded. Additionally, upon submission to NCBI's reference genome portal, a large number of very small duplicated scaffolds were detected, so we retained only one in each pair of duplicated sequences.

To compute descriptive statistics for the assembly we used the script *assemblathon_stats.pl* (https://github.com/KorfLab/Assemblathon, Bradnam et al., 2013), and genome completeness was assessed by assessing the percentage of complete single copy orthologues of the tetrapoda_odb10

database using *BUSCO* v5.3.2 (Simão et al., 2015). In addition, descriptive statistics were also computed for other lacertid genomes for comparative purposes (Kolora et al., 2018; Andrade et al., 2019; Yurchenko et al., 2020; rLacAgi1.pri: https://www.ncbi.nlm.nih.gov/refseq).

(c) Genome annotation

To annotate the assembly, we used the online semi-automated pipeline from *GenSAS* v6.0 (https://www.gensas.org, Humann et al., 2019), implementing a series of eukaryotic annotation resources. Initially, three libraries of predicted repeats were generated, one with *RepeatModeler* v2.0.1 (Flynn et al., 2020) with default settings and two with *RepeatMasker* v4.1.1 (Smit et al., 2013), using either the default GC content matrix and a 42-44% GC content matrix. The three libraries were combined to generate a consensus library of repeats to mask the genome assembly. Gene prediction was performed using *Augustus* v3.3.1 (Stanke et al., 2006) with default options and using the chicken (*Gallus gallus*) to model gene structure. Additional gene prediction runs were performed using software available in *GenSAS*, with non-default *Augustus* options and through a combination of several sources of evidence with *EvidenceModeler*. However, since the *Augustus* run with default setting obtained the highest *BUSCO* score of all approaches (see "Results"), this was retained as our gene set. Gene set refinement was performed with *PASA* v2.11.0 (Haas et al., 2008). Functional annotation of the consensus gene models was done via *BLAST* against the SwissProt database, with a BLOSUM62 matrix and remaining settings as default.

(d) Whole-genome resequencing

From each sample from our four main sites, genomic DNA was extracted from blood or tissue from tail tips using an EasySpin Genomic DNA Tissue Kit (Citomed), followed by a RNAse A digestion step. DNA purity and concentration was assessed by three different methods: electrophoresis with 1% agarose gel, spectrophotometry (Nanodrop) and fluorometric quantification (Qubit dsDNA BR Assay Kit, ThermoScientific). DNA from each individual was pooled by locality in equimolar concentrations for genomic library preparation. Individual paired-end libraries were generated using Illumina's Nextera XT Kit (Illumina, San Diego, IL, USA) following the manufacturer's protocol. The libraries were quantified by qPCR using the KAPA Library Quantification Kit and pooled. Sequencing was performed at Novogene UK in a NovaSeq instrument (Illumina) using 2x150bp reads.

(e) Read filtering and mapping

Quality control for the reads was evaluated using *FastQC* v0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc). *Trimmomatic* v.039 (Bolger et al. 2014) was used to remove adapter sequences, primers and low-quality reads with the following parameter settings: TRAILING:15, SLIDINGWINDOW:4:20 and MINLEN:30. Reads were then mapped to the *Gallotia galloti de novo* reference assembly with *BWA-MEM* (Li, 2013) using default settings. To maximise sequencing depth, all trimmed reads were mapped, including paired and unpaired reads. PCR duplicates were removed using *Picard* (http://broadinstitute.github.io/picard). Mapping statistics were calculated using *SAMtools* v.0.1.19 (Danecek et al., 2021) and custom scripts.

(f) Population genomics

To infer global population genetic parameters from the four populations, summary statistics were calculated based on allele count data to call single-nucleotide polymorphisms (SNP). To assess patterns of differentiation between populations, the fixation index (F_{ST}) was obtained using *PoPoolation2* v.1.201 (Kofler et al., 2011a) in non-overlapping windows of 100 kb, restricting to positions with a minimum coverage of 7, maximum coverage of 100, minimum count of the minor allele of 3 to consider a position as a SNP, and excluding windows for which less than 20% of the window met the coverage criteria. To calculate per-population levels and patterns of genetic variation calculated nucleotide diversity (π, Nei, 1987) and Tajima's *D* (Tajima, 1989) were calculated using *PoPoolation* v1.1.2 (Kofler et al., 2011b). Both statistics were calculated in non-overlapping windows of 100 kb, restricting to positions with a minimum coverage of 7, maximum coverage of 100, minimum count of the minor allele of 2 to consider a position as a SNP, and excluding windows for which less than 20% of the window met the coverage criteria.

(g) Selective sweep mapping

To detect regions of the genome potentially under selection in high-altitude populations, levels and patterns of genetic variation were compared along the genome. Given the main objective of comparing patterns between high and low-altitude populations, the aligned reads were merged from the three low-altitude pools to increase overall depth of coverage. To assess local patterns of genetic differentiation, F_{ST} was calculated in windows along the genome using *PoPoolation2*. Besides differentiation, selective sweeps are also likely to lead to localised distortions in the allele frequency

spectrum, such as reduction in genetic diversity and an excess of rare alleles. To test for this, *PoPoolation* provided the π and Tajima's D scores for both the high-altitude pool and the merged low-altitude pool. To highlight specifically selective sweeps acting in the high-altitude population, the mathematical formulae of π-ratio ($\pi_{\text{low-altitude}} / \pi_{\text{high-altitude}}$) and ΔD ($D_{\text{low-altitude}} - D_{\text{high-altitude}}$) applied by Gazda et al. (2018) were used in our work. By setting the comparisons this way, the results were polarised to highlight recent selective sweeps in high-altitude populations. In all these analyses, the following parameters were set to minimum base quality of 30, minimum coverage of 10, maximum coverage of 100, minimum count of the minor allele of 3 to consider a position as a SNP, and excluding windows for which less than 20% of the window met the coverage criteria. Window size was set to 50 kb, run in steps of 12.5 kb. Other windows sizes were also inspected (10 kb, 100 kb), but results were qualitatively unchanged, so they are omitted here. Scaffolds smaller than 500 kb were excluded from the analyses (the largest 54 scaffolds were used, accounting for 97.4% of the total assembly size).

The three statistics described above capture different properties of sequence data under selection, which means that combining them should increase the robustness of inference on selection. To combine the independent statistics, a de-correlated composite of multiple signals was implemented (DCMS; Ma et al., 2015). We used DCMS because it considers the correlational structure of the variables to weigh their relative contributions to the combined score. Following the procedure from Gazda et al. (2018) each value was first converted into fractional ranks to create uniformly distributed probabilities, to which an inverse-normal transformation was applied. Normalised scores were transformed into Z-scores, and from these *P*-values were calculated assuming a normal distribution of the data. Spearman's rank correlation coefficient was calculated for each pair of variables. Finally, correlation coefficients and *P*-values were used to calculate the DCMS score for each 50 kb window (taken from the previous analyses). We considered the top 0.1% of DCMS scores (as in Gazda et al., 2018) as stronger evidence of selection.

(h) Genotyping of candidate SNP using Sanger sequencing

To confirm the association between altitude and the two regions of the genome with the strongest signals of selection (in scaffolds 12 and 17), we selected for each one SNP for genotyping using Sanger sequencing. After identifying a SNP with high allele frequency differences between low and high-altitude populations we designed primers using Primer3Plus (Untergasser et al., 2007). We amplified the target region in scaffold12 using primers: F: 5'-TTTGGGTTTCCAGCATTAGC-3,

R: 5'-ATTCCCCCTGACTCCATTTT-3'; for scaffold17 we used primers: F: 5'-CGAGGTCCTAAGTGTGATCC-3', R: 5'-TTAATACATCCTTCAAGGCTTC-3.

PCR reactions were prepared with approximately 25 ng DNA, 5 μL of 2x Qiagen MasterMix, 0.4 μL of 10 pM of each primer and 3.2 μL of PCR-grade water, and ran under the following conditions: 1) an initial denaturing step of 95ºC for 15 min; 2) 5 touch-down cycles with 95ºC denaturing for 30 s, a 68-64ºC annealing temperature touchdown for 30 s and 72ºC extension temperature for 45 s; 3) 35 cycles with 95ºC denaturing for 30 s, a 64ºC annealing step for 30 s and 72ºC extension for 45 s; 4) a final extension at 60ºC during 20 min. After confirmation of amplification through 2% agarose gel electrophoresis, we prepared Sanger sequencing reactions, which were run on a 3500 XL Genetic Analyzer sequencer (Applied Biosystems). Sequencing files were inspected and aligned using *Geneious* (https://www.geneious.com). Spearman's correlations between allele frequencies and altitude were calculated with *PAST* v4.13 (Hammer et al., 2001).

(i) Functional annotation of variants

To explore the potential functional impact of variants within the regions identified as most likely under selection, we started by using *SAMtools* to extract reads from the high-altitude pool and the low-altitude merged pool for these specific regions (see Table 1, under Results), and used these reads to gather allele counts. These counts were used as input for the program *PoolSNP* (Kapun et al., 2020) to conduct localised variant calling, with the following parameters: minimum coverage of 10, maximum coverage of the 95% percentile of the distribution, minimum count of the alternative allele of 3, minimum frequency of 0.001, minimum base quality of 30 and discarding positions with missing data in one of the pools. Resulting VCF files were annotated using *SnpEff* v5.1 (Cingolani et al., 2012). We then used *PoPoolation2* to calculate differences in allele frequency for these annotated variants based on allele counts, using the following parameters: minimum base quality of 30, minimum coverage of 10, maximum coverage of 100 and minimum count of the minor allele of 3 to consider a position as a SNP. Detailed inspection of sequences mapping within candidate regions was performed using *IGV* v2.12.3 (Thorvaldsdóttir et al., 2013).

(j) Functional analysis of the BMPER *locus*

To characterize the *BMPER* locus in more detail, we started by extracting *BMPER* transcript sequences from 27 squamate species and chicken (since this species' gene model was used during

annotation) from NCBI's "Gene" portal. A single sequence was extracted from the following species (NCBI accession codes in parenthesis):

- *Podarcis muralis* (XM_028750015.1)
- *Podarcis raffonei* (XM_053409549.1)
- *Lacerta agilis* (XM_033165204.1)
- *Zootoca vivipara* (XM_035128736.2)
- *Rhineura floridana* (XM_061587150.1)
- *Anolis carolinensis* (XM_003222264.4)
- *Sceloporus undulatus* (XM_042476729.1)
- *Anolis sagrei ordinatus* (XM_060781627.1)
- *Pogona vitticeps* (XM_020788156.1)
- *Elgaria multicarinata webbii* (XM_063129831.1)
- *Varanus komodoensis* (XM_044419150.1)
- *Protobothrops mucrosquamatus* (XM_015811756.1)
- *Crotalus tigris* (XM_039336812.1)
- *Thamnophis sirtalis* (XM_014056655.1)
- *Thamnophis elegans* (XM_032234511.1)
- *Pantherophis guttatus* (XM_034408863.1)
- *Notechis scutatus* (XM_026674868.1)
- *Pseudonaja textilis* (XM_026702209.1)
- *Ahaetulla prasina* (XM_058182115.1)
- *Candoia aspera* (XM_063304049.1)
- *Python bivittatus* (XM_007426348.2)
- *Hemicordylus capensis* (XM_053264966.1)
- *Eublepharis macularius* (XM_054992126.1)
- *Sphaerodactylus townsendi* (XM_048511883.1)
- *Euleptes europaea* (XM_056857049.1)
- *Gekko japonicus* (XM_015405739.1)
- *Heteronotia binoei* (XM_060247726.1)
- *Gallus gallus* (XM_015281550.4)

These sequences, together with our annotated *G. galloti BMPER* coding sequence (2,538 bp, 14 exons), were imported into *AliView* v1.28 (Larsson, 2014), and aligned with *MUSCLE* v3.8.425 (Edgar, 2004), before manual inspection of alignments. The alignment identified exon 12 (881 bp) of the *G. galloti* sequence as a unique feature in this species across the species surveyed. To

evaluate the likelihood that this was a true exon, we modelled protein structure for the translated annotated coding sequence, as well as a translated sequence without this exon. Protein homology modelling was performed using the *SWISS-MODEL* web server (Guex et al., 2009; Waterhouse et al., 2018) by imputing the two translated amino acid sequences and running with default options. Model likelihood was assessed by comparing GMQE (global model quality estimate) and MolProbity scores.

To assess the functional impact of the missense mutation identified as a likely coding variant (T547A) we started by calculating GRAVY scores (grand average of hydropathy) for the "reference" and "alternative" amino acid sequences using *Sequence Manipulation Suite* v2 (Stothard, 2000). We further evaluated the impact of this amino acid change by modelling changes in protein stability (ΔΔG) using *DDMut* (Zhou et al., 2023), based on the model generated by *SWISS-MODEL* for the translated amino acid sequence derived from the coding sequence without the dubious exon 12.

References for extended methods

Andrade, P., Pinho, C., i de Lanuza, G. P., Afonso, S., Brejcha, J., Rubin, C. J., Wallerman, O., Pereira, P., Sabatino, S. J., Bellati, A., Pellitteri-Rosa, D., Bosakova, Z., Bunikis, I., Carretero, M. A., Feiner, N., Marsik, P., Paupério, F., Salvi, D., Soler, L., While, G. M., Uller, T., Font, E., Andersson & Carneiro, M. (2019). Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proceedings of the National Academy of Sciences*, 116(12), 5633-5642.

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

Bradnam, K. R., Fass, J. N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S., Chapman, J. A., Chapuis, G., Chikhi, R., Chitsaz, H., Chou, W.-C., Corbeil, J., Del Fabbro, C., Docking, T. R., Durbin, R., Earl, D., Emrich, S., Fedotov, P., Fonseca, N. A., et al. (2013). Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*, 2(1), 2047-217X.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80-92.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M. & Li, H. (2021). Twelve years of SAMtools and BCFtools. Gigascience, 10(2), giab008.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, *32*(5), 1792-1797.

Fernández-Rodríguez, I.; Barroso, F. M. & Carretero, M. A. (2021): An integrative analysis of the short-term effects of tail autotomy on thermoregulation and dehydration rates in wall lizards (*Podarcis bocage*i). *Journal of Thermal Biology*. 99: 102976.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, 117(17), 9451-9457.

García-Muñoz, E.; Ceacero, F.; Pedrajas, L.; Kaliontzopoulou, A. & Carretero, M. A. (2011): Tail tip removal for tissue sampling has no short-term effects on microhabitat selection by *Podarcis bocagei*, but induced autotomy does. *Acta Herpetologica*, 6(2): 223-227.

Gazda, M. A., Andrade, P., Afonso, S., Dilytė, J., Archer, J. P., Lopes, R. J., Faria, R. & Carneiro, M. (2018). Signatures of selection on standing genetic variation underlie athletic and navigational performance in racing pigeons. *Molecular Biology and Evolution*, 35(5), 1176-1189.

Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, *30*(S1), S162-S173.

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, R. & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology*, 9(1), 1-22.

Hammer, Ø.; Harper, D. A. & Ryan, P. D. (2001). Past: paleontological statistics software package for educaton and data anlysis. *Palaeontologia electronica*, *4*(1), 1.

Humann, J. L., Lee, T., Ficklin, S., & Main, D. (2019). Structural and functional annotation of eukaryotic genomes with GenSAS. In Gene prediction (pp. 29-51). Humana, New York, NY.

Kapun, M., Barrón, M. G., Staubach, F., Obbard, D. J., Wiberg, R. A. W., Vieira, J., Goubert, C., Rota-Stabelli, O., Kankare, M., Bogaerts-Márquez, M., Haudry, A., Waidele, L., Kozeretska, I., Pasyukova, E. G., Loeschcke, V., Pascual, M., Vieira, C. P., Serga, I., Monthchamp-Moreau, C. & González, J. (2020). Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. *Molecular Biology and Evolution*, 37(9), 2661-2678.

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C & Schlötterer, C. (2011a). PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PloS one*, *6*(1), e15925.

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

Kolora, S. R. R., Weigert, A., Saffari, A., Kehr, S., Walter Costa, M. B., Spröer, C., Indrischek, H., Chintalapati, M., Lohse, K., Doose, G., Overmann, J., Bunk, B., Bleidorn, C., Grimm-Seyfarth, A., Henle, K., Nowick, K., Faria, R., Stadler, P. F., & Schlegel, M. (2019). Divergent evolution in the genomes of closely related lacertids, *Lacerta viridis* and *L. bilineata*, and implications for speciation. *Gigascience*, 8(2), giy160.

Langkilde, T. and Shine, R. (2006). How much stress do researchers inflict on their study animals? A case study using a scincid lizard, *Eulamprus heatwolei*. *Journal of Experimental Biology*, 209, 1035-1043.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276-3278.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWAMEM. *ArXiv*:1303.3997 (2013).

Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H. (2015). Properties of different selection signature statistics and a new strategy for combining them. *Heredity* 115(5):426–436.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia university press.

Serén, N., Megía-Palma, R., Simčič, T., Krofel, M., Guarino, F. M., Pinho, C., ... & Carretero, M. A. (2023). Functional responses in a lizard along a 3.5-km altitudinal gradient. *Journal of Biogeography*. DOI: 10.1111/jbi.14711

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210-3212.

Smit AFA, Hubley R Green P. (2013). RepeatMasker Open-4.0. Available from: http://www.repeatmasker.org. Accessed 7th December 2021 through *GenSAS*.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic acids research*, *34*(suppl_2), W435-W439.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, *28*(6), 1102-1104.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585-595.

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

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic acids research*, *35*(suppl_2), W71-W74.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, *46*(W1), W296-W303.

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome research*, 27(5), 757-767.

Yurchenko, A. A., Recknagel, H., & Elmer, K. R. (2020). Chromosome-level assembly of the common lizard (*Zootoca vivipara*) genome. *Genome biology and evolution*, 12(11), 1953-1960

Zhou, Y., Pan, Q., Pires, D. E., Rodrigues, C. H., & Ascher, D. B. (2023). DDMut: predicting effects of mutations on protein stability using deep learning. *Nucleic Acids Research*, *51*(W1), W122-W128.

Supplementary figures

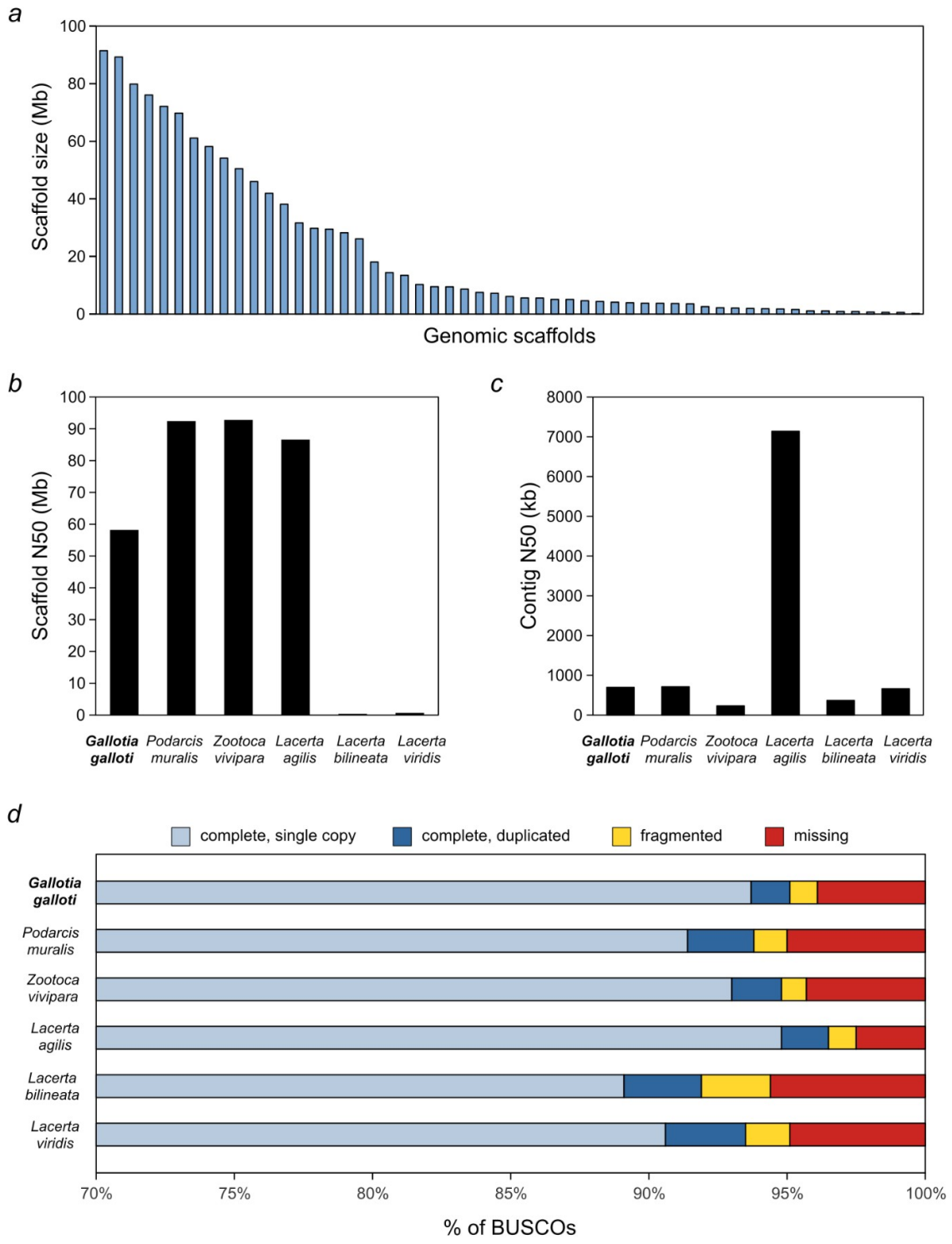

Figure S1. Comparison between the *Gallotia galloti de novo* assembly and other published lacertid genomes (*Lacerta agilis*, *Lacerta bilineata*, *Lacerta viridis*, *Podarcis muralis* and *Zootoca vivipara*). (a) Size (Mb) of the top 55 genomic scaffolds of *G. galloti* (all scaffolds larger than 100 kb). (b) Scaffold N50. (c) Contig N50. (d) *BUSCO* scores (aligned to the tetrapoda_odb10 database).

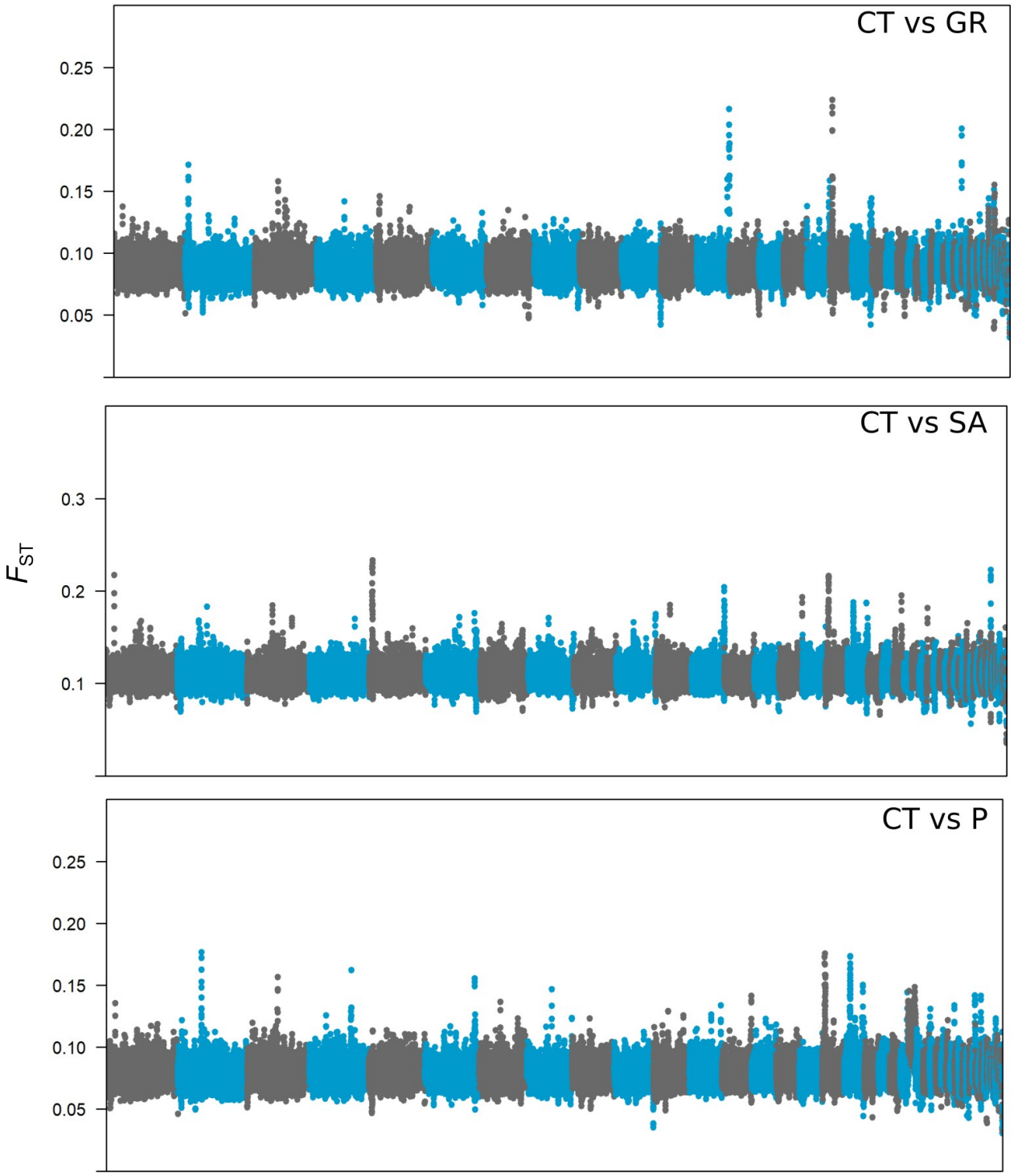

Figure S2. Genetic differentiation (fixation index, F_{ST}) between high-altitude (CT) and low-altitude (GR, SA, P) populations of *G. galloti*. Each dot corresponds to a 50 kb window, with successive windows having a 12.5 kb overlap. Alternating colours indicate the different genomic scaffolds.

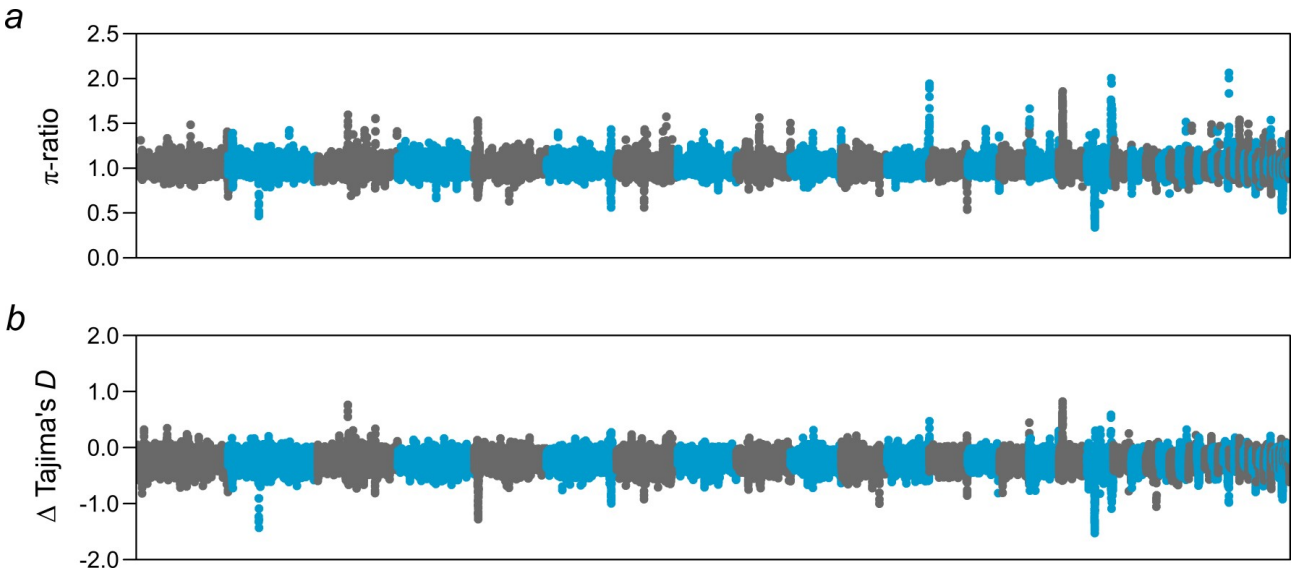

Figure S3. Genome-wide signatures of selection potentially associated with high-altitude adaptation in *Gallotia galloti*. (a) Ratio between the nucleotide diversity (π, %) of low-altitude and high-altitude populations. (b) Differences in the allele frequency spectrum (deviations from neutrality) of low and high-altitude populations as calculated with Tajima's *D*. By setting the comparisons this way, the results were polarised to highlight recent selective sweeps in high-altitude populations.

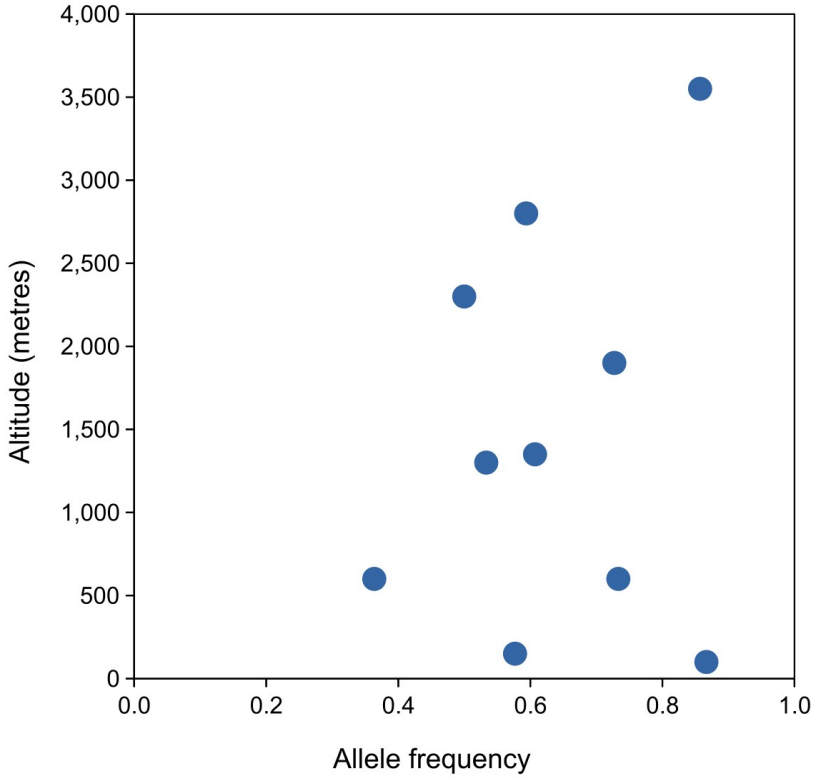

Figure S4. Allele frequencies for a single-nucleotide polymorphism from scaffold17, genotyped using Sanger sequencing (n = 143; 10 populations). Despite a large variability between populations, which may explain the signal that was detected in the genome-wide scan for signatures of selection, we did not detect a strong association with increasing altitude (Spearman's correlation: $\rho = -0.024$, $P = 0.947$).

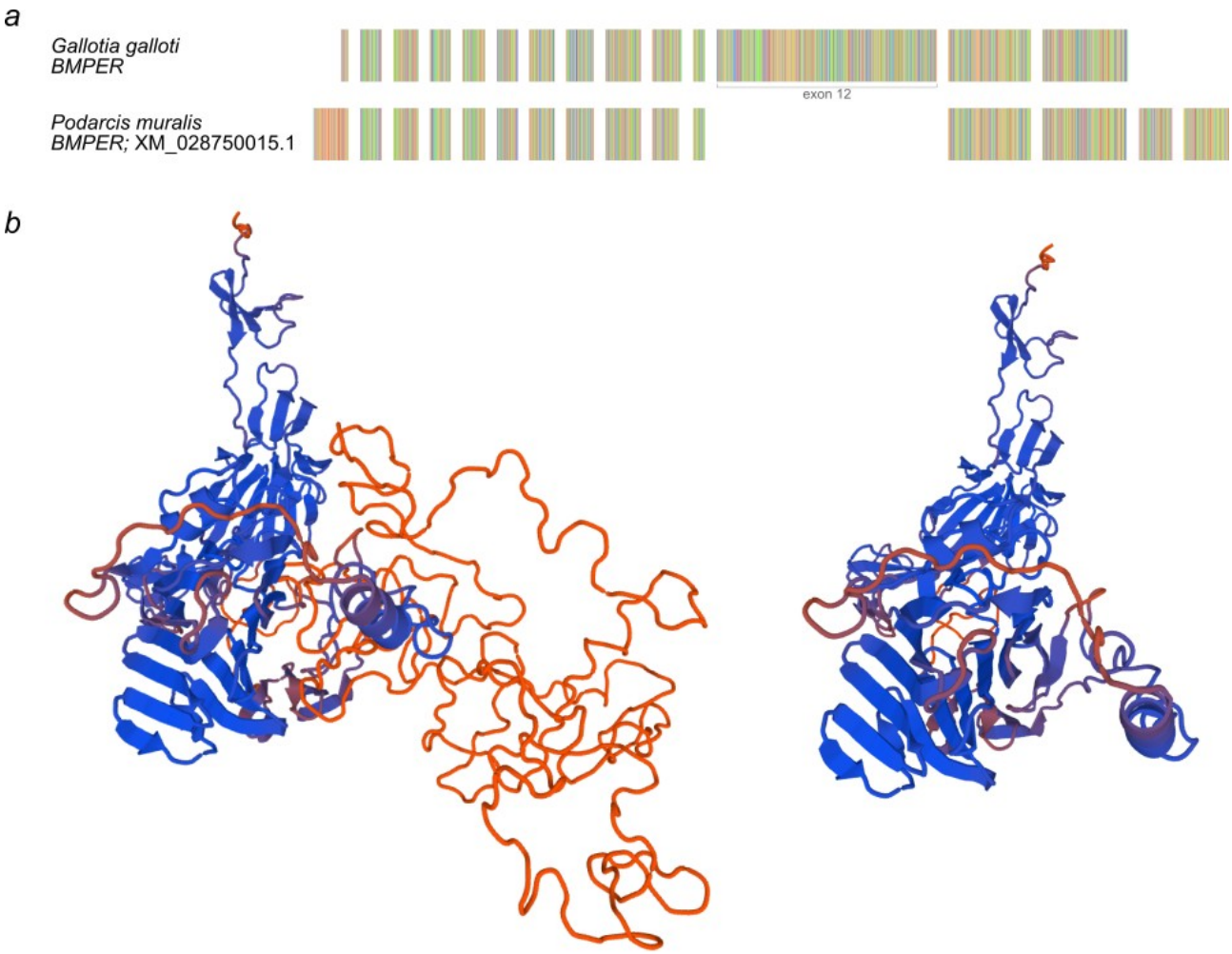

Figure S5. Exon structure of the *BMPER* gene (scaffold12) in *G. galloti*. (a) Alignment of our annotated coding sequence of *G. galloti BMPER*, compared to the coding sequence of the same gene for the lacertid *Podarcis muralis* (NCBI, XM_028750015.1). The 12th exon of the *G. galloti* annotation is highlighted. Colours in the alignments correspond to different nucleotides. (b) Structural models for the BMPER protein, generated by homology modelling with SWISS-MODEL, using as template from a corn snake (*Pantherophis guttatus*) BMPER prediction (A0A6P9B210.1.A isoform X4, generated by AlphaFold). The left model corresponds to the best model based on the amino acid sequence translated from the full annotated *BMPER* CODING sequence. The model on the right corresponds to the best model based on the amino acid sequence translated from the annotated *BMPER* coding sequence, minus the exon 12 derived amino acids. Models are coloured based on model-template alignment confidence (red – lower confidence; blue – higher confidence).

Supplementary tables

Table S1. List of populations and number of individuals (all males) sampled for this study. Populations with an asterisk (*) were used for whole-genome pool-sequencing (see Table S3).

Altitude (metres a.s.l.)	Location	Approximate coordinates	n
100	El Médano (EM)	28.060,-16.545	16
150	Poris de Abona (P)*	28.165,-16.450	14
600	Granadilla de Abona (GR)*	28.121,-16.573	18
600	La Sabinita (SA)*	28.183,-16.499	12
1300	Área Recreativa (AR)	28.209,-16.540	15
1350	Vilaflor (VL)	28.150,-16.632	16
1900	Corona Forestal (CF)	28.234,-16.540	11
2300	Base of Teide (BT)	28.255,-16.621	14
2,750-2,900	Teide refugia (T)	28.271,-16.618	16
3550	Cone of Teide (CT)*	28.271,-16.640	15

Table S2. Summary statistics of the *de novo* reference genome assembly and annotation of *Gallotia galloti*.

	CIBIO_Ggalloti_1.0
Genome size (Gb)	1.18
Scaffold N50 (Mb)	58.15
Contig N50 (kb)	698.54
Number of scaffolds	7,039
Number of scaffolds >1 Mb	49
Number of scaffolds >100 kb	57
% GC	43.14
Complete single-copy BUSCOs (%)	7,006 (93.7%)
Complete duplicated BUSCOs (%)	103 (1.4%)
Fragmented BUSCOs (%)	74 (1.0%)
Missing BUSCOs (%)	297 (3.9%)

Table S3. Summary statistics of the whole-genome pool-sequencing dataset.

Population	Number of individuals in pool	Total number of reads	Percentage of duplicate reads	Coverage after duplicate removal	Percentage of mapped reads
CT (Cone of Teide)	14	285,764,025	67.0%	21.1	98.60%
GR (Granadilla de Abona)	16	113,667,347	50.5%	12.5	99.20%
SA (La Sabinita)	12	91,542,222	54.7%	9.2	99.00%
P (Porís de Abona)	12	141,685,537	51.9%	15.5	99.14%

Table S4. Genomic intervals with evidence of selection to high-altitude in *Gallotia galloti*, highlighted from a composite selection statistic (DCMS) and ordered from highest to lowest DCMS value. Protein coding genes in each interval were retrieved from the *G. galloti* genome annotation, considering a 100 kb buffer either side of the genomic interval with evidence of selection; gene names follow their human orthologues. Protein coding sequences with no known human orthologues (marked with *) were given a specific code, following the annotation.

Scaffold number	Coordinate interval (kb)	Coordinate range (kb)	Genes within or overlapping an additional 100kb in each direction from the interval	Top DCMS value (50kb window)
17	3,425,000-4,062,500	637.5	*PPP2R2A, BNIP3L, DPYSL2, ADRA1A, Gal.00g130040*, Gal.00g130060**	23.36
12	41,637,500-41,895,811	258.3	*BMPER, FANCD2OS, NPSR1*	21
3	29,875,000-29,962,500	87.5	*RPS6KA3*, *CNKSR2*, *BCOR*	19.48
15	29,625,000-29,712,500	87.5	*TLN2*	19.4
3	57,362,500-57,450,000	87.5	-	18.72
18	23,575,000-23,675,000	100	*DNM1*, *CIZ1*, *BBLN*, *PTGES2*, *ODF2*, *ABCA2*, *FUT7*, *NPDC1*	18.72
30	5,437,500-5,533,213	95.7	*TPBG, Gal.00g165360*, C7orf25*	18.31
16	362,500-437,500	75	*SLC25A3, DMRTB1, GLIS1*, *NDC1*	17.13
16	29,287,500-29,450,000	162.5	*ASTN1*, *PAPP2*, *PEG10*, *TECR*	16.56
7	49,537,500-49,612,500	75	*RERE*, *ENO1*, *CA6*, *SLC2A5*	15.72
40	3,350,000-3,462,500	112.5	*COP1*	15.11
6	63,725,000-63,775,000	50	*SPG16*	14.21
11	18,437,500-18,487,500	50	*LRIG2*, *MAGI3*	14.09