Title Page

Title: Molecular insights into high altitude adaption and acclimatisation of *Aporrectodea caliginosa*

Running Title: *A. caliginosa* a*daption & acclimatisation*

Author list: Iain Perry[1,2,*], Szabolcs Balazs Hernadi[1], Luis Cunha[3,4], Stephen Short[1,5], Angela Marchbank[6], David Spurgeon[5], Pablo Orozco-terWengel[1], Peter Kille[1].

[1] Organisms and Environment, Cardiff University, Wales, CF10 3AX, UK.

[2] Wales Gene Park, Cardiff University, Wales, CF14 4XN

[3] Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal.

[4] School of Applied Sciences, University of South Wales, Pontypridd, Wales, CF37 4BD, UK.

[5] UK Centre for Ecology & Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire England, OX10 8BB, UK.

[6] Genomics Hub, Cardiff University, Wales, CF10 3AX, UK

***Address for correspondence:** perryia3@cardiff.ac.uk

Keywords: High-altitude, Adaption, Acclimatization, *Aporrectodea caliginosa*

Summary blurb: A Megabase genome assembly for *A. caliginosa* is presented with transcriptomic and SNP based evidence for acclimatisation and adaption to extreme weather conditions found at high altitude.

Main text

Abstract

Here we explore the high-altitude adaptions and acclimatisation of *Aporrectodea caliginosa*. Population diversity is assessed through mitochondrial barcoding, identifying closely related populations across the island of Pico (Azores). We present the first Megabase N50 assembly size (1.2 Mbp) genome for *A. caliginosa* is presented. High and low -altitude populations were exposed experimentally to a range of Oxygen and temperature conditions, simulating altitudinal conditions and the transcriptomic responses explored. SNP densities are assessed to identify signatures of selective pressure and their link to differentially expressed genes. The high-altitude *A. caliginosa* population had lower differential expression and fewer co-expressed genes between conditions indicating a more condition refined epigenetic response. Genes identified as under adaptive pressure through F_{st} and nucleotide diversity in the high-altitude population clustered around the differentially expressed upstream environmental response control gene, HMGB1. The high-altitude population of *A. caliginosa* indicated adaption and acclimatisation to high altitude conditions and suggested resilience to extreme weather events. This mechanistic understanding could help offer a strategy in further identifying other species capable of maintaining soil fertility in extreme environments.

Introduction

Earthworms provide a critical role as terrestrial eco-engineers in the recycling of biological material, improving soil fertility and are commonly used as part of environmental management to improve poor soils (Butt 2010; Blouin et al. 2013). The higher temperatures and increased extreme weather events associated with global warming has increased the frequency of crop losses (Powell and Reinhard 2016). This has led to some crops being increasing planted at higher altitudes where cooler conditions can be found, elevating the importance of understanding mountain soil fertility (Lenoir et al. 2008; Skarbø and VanderMolen 2015). Significant investment has been focused on improving the resilience of cultivated crops, as well as developing strategies to manage exposure to extreme weather events at low land and facilitate crop cultivation on mountainsides (Gale 2004; Zorn et al. 2005; Butt and Briones 2011; Skarbø and VanderMolen 2015; Tolessa et al. 2017). Earthworm's play a critical role in decomposition, nutrient cycling, and crop yield, role in crop yield, the impact of climate change on earthworm communities has been recently comprehensively been reviewed (Singh et al. 2019). However, this review has highlighted the limited consideration that has been afforded to understanding the resilience of annelid species at high altitudes that are critical for soil fertility, a resilience vital to their persistence at such locations.

High altitude mountains at temperate latitudes present extreme environmental conditions with large temperature fluctuations and changes in oxygen availability as atmospheric pressures decrease with altitude (Peacock 1998). While vagile species may cope with such environmental stressors by changing their location, less mobile species must adapt or acclimatise to those conditions to avoid perishing. As ectotherms earthworms are vulnerable to temperature changes (Costanzo and Lee 2013), and while some anecic species use deep burrows to avoid temperature extremes, shallow burrowing endogeic and epigeic species are particularly susceptible. Adaption and acclimatisation of humans and other mammals to high altitude, and in particular hypoxia, has been an area of significant research (Manalo et al. 2005; Ke and Costa 2006; Majmundar et al. 2010; Scheinfeldt and

69 Tishkoff 2010; Storz et al. 2010), but few studies have attempted to investigate altitudinal impacts
70 on earthworms (Gonzalez et al. 2007; Kanchilakshmi 2016) Earthworms rely upon a relatively
71 unique system that uses cutaneous respiration and Erythrocruorin to deliver oxygen to its tissues
72 (Giardina et al. 1975; Royer Jr et al. 2006), and mechanisms of response to atmospheric oxygen
73 deprivation for this system are not well described.
74
75 To investigate adaptive and acclimatised responses of earthworms to high altitude, we can look at
76 differential expression of genes and at variation in the distribution of single nucleotide
77 polymorphisms (SNPs) across the genome that may be indicative of selective processes (Robledo et
78 al. 2017). However, while such analyses are readily feasible for many species that have had their
79 genomes assembled, earthworms, as a group, are underrepresented among these with *Eisenia fetida*
80 (N50 <10 Kb), *Eisenia andrei* (N50 ~740 Kb), *Metaphire vulgaris* (N50 ~4.2 Mb) and *Amynthas corticis*
81 (N50 ~31 Mb) representing the few earthworms with their genomes sequenced (Zwarycz et al.
82 2015; Bhambri et al. 2017; Jin et al. 2020; Shao et al. 2020; Wang et al. 2021). Assembling a high-
83 quality genome (N50 >10 Kbp) for an earthworm has proved a challenge for many years due to their
84 high allelic diversity (Rimington 2022), however, with long read technology combined with
85 scaffolding techniques, it is possible to now overcome this hurdle (Ghurye and Pop 2019).
86 A fortuitous combination of geography and human migration offers us the chance to investigate
87 molecular mechanisms underlying the capacity of earthworms to adapt to high altitudes. The
88 geologically young Azores archipelagos began emerging a little over 8 million years ago, but the
89 youngest island. Pico, volcanically grew only around 270 thousand years ago (Miranda et al. 1998;
90 Carine and Schaefer 2009) with the main volcanic peak of Pico erupting only 300 years ago
91 (Woodhall 1974). Estimate of human settlement with farming range ca. 15th century to potentially as
92 far back as the 11th Century (Pacheco et al. 2010; Rodrigues et al. 2015). It is likely that many of the
93 species, including earthworms, on the island are semi-invasive, brought over with human travel from
94 multiple locations (Novo et al. 2015). Even with the earliest of potential earthworm stowaways,

there has been very little time for species to adapt to the island, and less time still to adapt to high altitude following eruptions.

In this study, we develop the first genome for *A. caliginosa* and generate a masked and annotated genome. We then go on to use this new resource in conjunction with differential expression, F_{st} and nucleotide diversity (Pi) to help to identify the mechanisms by *A. caliginosa* has adapted or acclimatised to extreme climate conditions found at high altitudes as derived from populations resident to low and high altitudes in the Island of Pico within the Azorian archipelago.

Results

Species and population selection

Of the 165 *A. caliginosa* individuals collected only for COII analysis (Figure 1A), for 619 nucleotide sites, Theta-W was 0.00427, Pi 0.00562 and Tajima's D 0.6265 (P>0.10). These suggest a low level of genetic diversity across the island of Pico and when combined with Baeyesian skyline analysis, indicates that the seeding population had undergone a long-term population contraction since the last glacial maximum (Figure 1B).

The origins of Pico *A. caliginosa* was assessed through generation of haplotype groups and calculating a time rooted phylogeny compared to Western Palearctic *A. caliginosa* individuals (Figure 1C). A distinct linage split between northern Spain and the French Pyrenees *A. caliginosa* individuals was detected, diverging between 27.9 and 107.4 million years ago. However, individuals from Pico clustered into multiple haplotypes found in several areas across the Western Palearctic including Ireland, France, Austria, and Southern Spain potentially indicating multiple introduction events.

Genome generation

More than 19.5 µg high molecular weight pure DNA was extracted from a single *A. caliginosa* individual from Pico (2200 m asl). Sequencing generated a total of: 24.59 Gbp of long read (~10 Kbp) Nanopore reads (from three runs), 97 Gbp of 10X Chromium barcoded PE150 reads and 60 Gbp of PE150 short read data. An initial assembly of paired-end error corrected Nanopore reads with Wtdbg2 generated an N50 of 159 Kbp. This was successively increased to an N50 of 1.2 Mbp with scaffolding from 10X Chromium reads using Nanochrome. BUSCO indicated 93.4% of core metazoan genes are present. Genome

annotation identified 42,566 gene objects across the masked genome, relating to 25,556 genes were identified (Figure 2).

Differential expression analysis

Between 4 and 11 million reads were sequenced for each individual, with mapping statistics to the genome similar between both high and low altitude population individuals. When all samples were analysed together through DESeq2, and principal component analysis run, principal component 2 split the expression profile of native individuals with experimental individuals, while Principal component 3 clearly separates HA and LA populations (Figure 3A).

The LA population had a larger number of genes showing differential expression between treatment pairs than the HA population. Differentially expressed genes for each population in each condition were assessed via DiVenn. This indicated that many of the differentially expressed genes between treatments for the LA population overlapped indicating a shared response, while the HA population indicated a more nuanced response with a reduced level of differential expression between conditions (Figure 3B, 3C, 3D). Of particular note, the environmental response High-Mobility group (HMG) genes, HMGB1 and HMGB2, were universally upregulated with a Padj <0.05 in both HA and LA populations in all temperature comparisons of 4°C Vs 21°C (with fold changes for both populations ranging between 1.8 and 3.2 for HMGB1 and 1.8 and 3.3 for HMGB2). Exposure to different temperatures appear to have the greater impact on differential gene expression than a reduction in atmospheric oxygen.

SNP analysis

A phylogeny of individuals assessed in the differential gene expression analysis and SNP analysis was calculated from their SNPs across the genome. There was a clear difference in SNP compliments between HA and LA individuals, indicating a rapid accumulation of SNPs separating the HA population from the LA population (Figure 4A). The LA native individuals taken from a separate LA native fauna site indicate a slight divergence from those collected for LA experimental exposures.

The PCA and MDS plots (Figure 4B, 4C) generated from the SNPs support the divergence observed in this phylogeny, while the Manhattan plot indicates the number of genomic regions under selective pressure (Figure 4D). A total of 358 Genes passed filters for F_{st} and nucleotide diversity (Pi) (Figure 4E). A further filter was applied to remove genes with fewer than 5 SNPs per gene leaving 271 genes identified as under selective pressure.

To identify if any of the 271 genes had direct interactions with each other or to the genes HMGB1 and HMGB2 that were identified as universally upregulated in HA and LA differential expression, interactions were examined in STRING (Figure 5A). This analysis identified that while HMGB1 is not itself under selective pressure, its interaction with TP53 forms the epicentre of a network of genes that are. Through putative transcription factor enrichment analysis and upstream expression to kinase analysis, upstream control of SNP associated genes were identified in differentially expressed genes (Figure 5B) and the overlap between all sets of identified genes are identified (Figure 5C) with SOX2 appearing in all four gene lists.

Discussion

To explore the adaptions and acclimatisation of earthworms to altitude, a population of *A. caliginosa* was identified on the strato-volcanic island of Pico, Azores. The population introduced to the island demonstrates the historical bottlenecking of diversity from the last glacial maximum, but no further expansion or contraction is detected. The date of introduction of *A. caliginosa* to the island has not yet been identified, however the evidence of genetic bottlenecking suggests this is no older than 10,000 years ago and more probably occurred with the arrival of human settlers with transplanted crops in the last millennium (Ashe 1813; Pacheoco et al. 2010; Rodrigues et al. 2015). All measures of mitochondrial diversity indicated a low genetic diversity between high and low altitude populations. Critically, this suggested populations could be used effectively, and operate close to the model inbred mouse strains that are used in mammalian research.

The genome generated as a tool for investigating differential expression and SNPs, is the first genome for the species *A. caliginosa* and is one of the first megabase assemblies for earthworms. This presents a significant advance in the resources for one of the most cosmopolitan and widely distributed earthworm species, that is used heavily in environmental monitoring and for research (Bart et al. 2018). The use of Nanochrome software greatly improved the assembly of scaffolds, with some contigs now close to chromosome arm length. The assembly demonstrated high completeness as assessed via BUSCO and purity of assembly as assessed via Blobtools. The megabase assembly is one of the most contiguous assemblies for an earthworm assembly by a considerable size and leave a short step to chromosome level assembly. The future use of Chromosome

conformation capture (Hi-C) should allow the final scaffolding of a full diploid chromosomal assembly from the scaffolds present, overcoming the very high levels of haploid diversity.

In replicating high altitude conditions, exposures to low temperatures (having spent 6 months at 21°C) induced more differential gene expression than reduction in oxygen concentrations. This is not unexpected as the cutaneous respiration of earthworms and their use of Erythrocruorin avoids the well explored pressures of respiration in mammals (Giardina et al. 1975; Royer Jr et al. 2006; Storz 2016). With increasing adverse weather events earthworms may be exposed to lowered levels of oxygen on a more regular basis e.g. fluctuations in soil saturation from heavy rain that can drive worms from burrows to surface. For many populations these conditions are temporary, however earthworms have been found to survive in other extreme environments with very low levels of oxygen and high levels of carbon dioxide (volcanic caldera) (Cunha et al. 2011). Similarly, for those identified at high altitude there is no temporary escape to a more oxygen rich surface. Despite the acclimatisation of LA and HA earthworm populations in identical conditions for 6 months, a timeframe that should have reduced interference from native differential expression of genes that are not determined by hereditary factor, the LA population still had a significantly higher level of differentially regulated genes to both temperature and oxygen availability than the HA population. As indicated in PCA analysis, where native expression of HA and LA populations group with experimental HA and LA populations, response to the simulated climatic conditions suggests that populations' responses are rooting from either adaptive changes, or from epigenetics that prime for rapid environmental response beyond the 6 month pre incubation. In the HA population, fewer genes show differential expression between paired temperature comparisons, potentially indicating a more nuanced condition

221 specific response to survival in extreme weather. The HA population's adaption through
222 transcriptome remodelling could be because their existing physiological systems were more
223 adaptable. The high levels of differential expression in the LA population suggest a
224 widespread adaptive change founded in a stress like response. Thus, in the LA population, a
225 number of common stress response pathways such as ERK/MAPK and Wnt signalling were
226 differential expressed, indicating a rapid modulation of protective stress pathways that are
227 not required in the HA population. The identification of HMGB1 and HMGB2 as universally
228 upregulated is of particular note. The primary function of these two genes are in the
229 regulation of environmental response, operating as a DNA chaperone for pathways involved
230 in (but not limited to): inflammation, immune response, DNA repair and hypoxia response
231 (Muller et al. 2004; Kang et al. 2014). These genes act as chromatin-binding factors or,
232 through extracellular release, binding to RAGE, TLRs and activating MAPK and NFκB (Bianchi
233 2009). Despite both populations displaying upregulation of HMGB1 and HMGB2 in response
234 to the temperature shift, the lower levels of differentially expressed genes within the HA
235 population suggests that the HA population might be utilising a separate response
236 mechanism not used in the LA population.
237
238 SNP analysis identified a large number of genes as under selective pressure in the HA
239 population compared to the LA population. Many of these genes are responsible for
240 proteins that form interacting networks that HMGB1 critically forms the epicentre of. That
241 network involves multiple transcription factors and cell cycle regulators such as NFκB1
242 which was identified as under selective pressure and directly interacts with HMGB1 itself
243 has a wide variety of biological interactions, particularly in stress response (Concetti and
244 Wilson 2018). HMGB1's role in acclimatory thermal acclimatisation has been previously

245 explored in detail including their role as 'general activators' rather than solely specific gene
246 regulators (Somero 2005). HMGB1 functions to allow increased gene-specific transcription
247 factors. While in the LA population this explains the high levels of differential gene
248 expression the pattern is not repeated in the HA population. It is possible that the lower
249 levels of differential gene expression seen in the HA population is because of some or a
250 combination of the changes identified in the networks of interacting genes, changes to
251 upstream promoters outside of coding regions sequenced in this study or epigenetic
252 modification. These changes minimise energy expenditure to these stressors and are
253 providing long-term adaption to the extreme weather shifts seen at high altitude.
254
255 The significant limitation of the current study is the absence of fitness data which prevents
256 us from determining comparative phenotypic adaptation between the two populations.
257 However, the genetic changes identified are indicative of adaption and acclimatisation of
258 the HA population which is particularly remarkable considering the last volcanic eruption
259 300 years ago is likely to have erased any existing populations, meaning the current HA
260 population represents a LA population that has returned and rapidly adapted and
261 acclimatised (Woodhall 1974). Earthworms are not alone in rapid island evolution, as
262 morphological evolution has been observed in several families and taxa including mammals,
263 lizards, crabs (Schubart et al. 1998; Millien 2006; Eloy de Amorim et al. 2017), though few
264 studies have attempted to understand the genetics behind the observed morphological
265 changes in even isolated mainland environments (Lyman et al. 2002).
266
267 **Conclusions**

The population of *A. caliginosa* from Pico, Azores, indicates a putative epigenetic acclimatisation and adaption to environmental stressors. In particular, the HA population appears capable of rapid acclimatisation to changes in extreme weather events without the need for transcriptional modification of large numbers of genes. This opens the prospect of their ability to survive increasingly common extreme weather events associated with human accelerated global warming climatic change. High altitude environments could prove critical population either for reseeding resilient populations following localised extinction events or for the engineering of improved fertility HA soils for growing crops at HA. The identification of other species that are similarly endowed are important areas for further investigation.

Methods

Species and population selection

A. caliginosa individuals from across the island of Pico (Table S1), Azores (38.467N 28.400W), were collected in 96% EtOH. DNA extraction and purification were carried out with Qiagen DNeasy Blood and Tissue Kit (Qiagen, UK). Samples were quantified via NanoDrop (ND-1000) prior amplification of a fragment of the mitochondrial DNA COII gene (Pérez-Losada et al. 2009). The PCR amplification of the COII fragment consisted of an initial DNA denaturation at 95°C for 5 minutes followed by 36 cycles of a) DNA denaturation at 95°C for 1 minute, b) annealing at 50°C, and c) extension of PCR products at 72°C for 1 minute, ending with a final extension at 72°C for ten minutes. PCR reactions consisted of a 25 µL reaction volumes containing: 5x PCR Buffer Flexi Green, 2.5 mM $MgCl_2$, 10 mM dNTPs, 1 U Taq polymerase (Promega, UK), 10 pM COII primers (Pérez-Losada et al. 2009) and ~ 50 ng of template DNA. PCR products were visualised via Qiaxcel (Qiagen, UK) and sequenced by Eurofins Genomics (Germany). Sequences were quality assessed before aligning and

trimming to the same length (619 bp) with MEGA (v7.0.26) (Kumar et al. 2016). Theta-W, Pi and Tajima's D were calculated using DnaSP for the high and low populations separately (Rozas et al. 2017).

A time calibrated Bayesian phylogenetic tree was calculated using BEAUti and BEAST (v1.8.4) with an estimated molecular divergence rate of 2.4% per site per million years (Chang and James 2011). BEAST was run independently ten times with the Markov Chain Monte Carlo algorithm run for 20 million steps and discarding the initial 25% of the steps as burn-in. After removing the burn-in the remaining of the runs were merged using LogCombiner. Convergence of the MCMC were assessed using Tracer (v1.6) (Drummond et al. 2012), with the later also used to estimate the haplotype network.

High molecular weight DNA isolation

Muscular tissue from a high altitude (HA) individual from Pico was dissected (avoiding the clitellum and 5 segments from the tail tip) and digested for 6 hours in 1080 µL ATL buffer, 40 µL Proteinase K at 20 mg/mL, and 20 µg/mL RNase (Qiagen) at 56°C. DNA was purified via Phenol/Chloroform extraction and precipitated in isopropanol containing 0.2% ammonium acetate prior to hooking with a glass rod. The isolated DNA was then washed in 70% ethanol and dissolved in elution buffer (Qiagen) for 48 hours. The resulting high molecular weight (HMW) DNA (fragments >40 Kbp) was assessed via Nanodrop, Qubit and agarose gel electrophoresis.

Genome sequencing and assembly

Fifty-fold coverage of barcoded 10X Chromium sequenced PE150 and 50 fold coverage of unbarcoded short read sequencing PE150 was performed by Novogene. For Nanopore

316 sequencing, 5 µg of HMW DNA was fragmented to ~16 Kbp in a Covaris g-Tube before SPRI
317 bead clean-up (Beckman-Coulter IN, USA) to remove small molecular weight fragments
318 below ~3000 Kbp. The fragmented DNA was sequenced on three Oxford nanopore Minion
319 flow cells (MK1 R9 RevD) following manufacturer's protocols.
320
321 Combined Nanopore long read data was error corrected using the short read data with
322 FMLRC (Wang et al. 2018) prior to assembly with Wtdbg2 (Ruan and Li 2020). The assembly
323 was further improved with transcriptomic data for *A. caliginosa* using L_RNA scaffolder (Xue
324 et al. 2013). The large contigs were further scaffolded with 3 rounds of Nanochrome that
325 uses barcoded 10x Chromium reads supported by Nanopore long read data (Rimmington
326 2019). Between each round LR_Gapcloser and SOAP gap closing was run to fill gaps
327 identified in Nanochrome scaffolding (Xu et al. 2018).
328
329 BUSCO v1.0, Blobplots v1.0 and contig statistics were used to assess the quality of the
330 assembly and contaminating reads from the earthworm symbiote *Verminephrobacter* were
331 removed (Laetsch and Blaxter 2017; Waterhouse et al. 2017). Repeat masking of the
332 genome assembly was performed with RepeatModeler and RepeatMasker (Smit et al.
333 2015). The masked assembly was used to generate an annotations file with OmicsBox,
334 Blast2GO and Interproscan (Biobam). Gene sequences in the masked genome were
335 annotated by performing a BlastP search (E-value cut off 10^{-5}) against the human proteome
336 (UP000005640).
337
338 <u>Experimental setup, exposures, RNA extraction, Library generation and sequencing</u>

Populations (n200 individuals) of earthworms were collected from Pico at high altitude (HA) (38.46447N 28.40615W, 2073 m asl) and low altitudes (LA) (38.49733N 28.26995W, 196 m asl). HA and LA 'native' individuals were also collected in RNAlater (ThermoFisher, UK) for native RNA expression analysis. HA native individuals were collected from the same site as live HA individuals for experimental work. LA native individuals were taken from a LA site close (at a similar altitude) to the LA experimental individuals where endemic flora was present but where sufficient experimental LA worms could not be collected (38.48672N 28.25272W, 340 m asl).

Analysis of 'native' individuals provide insight into impact of chronic exposure to the climatic difference that exist at high altitude. Experimental exposures were design probed the acute response indicative of short-term weather patterns. Prior to experimental exposures, HA and LA experimental populations were kept for 6 months to acclimatise to the same conditions, i.e., room temperature in 20 L boxes of soil comprised of equal ratio of topsoil, compost, and bark chipping. Soil was kept moist, and earthworms were supplied with 0.5 Kg horse manure as feed every 2 months. For the experimental study, earthworms were placed in identical containers (20 cm x 14 cm) with mesh lids in 1 kg of soil and exposed to various temperature and oxygen for two weeks to replicate different climatic conditions. Conditions selected cover a factorial combination of high and low temperature and three oxygen levels, specifically, 21°C at 13% O_2, 21°C at 16% O_2, 21°C at 20% O_2 and 4°C at 13% O_2, 4°C at 16% O_2, 4°C at 20% O_2. HA and LA populations were exposed to each condition with 10 worms from each population per container. Worms were washed and weighed before and after exposure. Temperature was controlled in an Innova 4230 refrigerated

362 incubator (New Brunswick Scientific) and oxygen with a ProOx model C2` (BioShperix, NY,
363 USA).
364
365 Immediately following the exposure, earthworms were hand sorted from the soil and RNA
366 extracted from a 1 cm posterior section (5 segments from tail tip, avoiding any clitellum
367 tissue). Tissue was homogenised in 600 µL TRIzol (ThermoFisher, UK) before RNA was
368 purified with Direct-Zol RNA MiniPrep kit (Zymo Research, CA, USA). Purified samples were
369 quantified by HS RNA Qubit (Invitrogen), and RNA quality assessed with a Qiaxcel RNA
370 cartridge (Qiagen). Three samples from the high and low altitude populations for each
371 exposure condition, were selected for sequencing, with individuals with the highest RNA
372 concentration selected for the RNASeq library preparation. Libraries were generated with a
373 KAPA mRNA HyperPrep kit (Roche, CA, USA). Samples were assessed via a D1000 chip on an
374 Agilent Tapestation and samples pooled evenly at 1nM. All 42 samples were sequenced
375 twice 1x 75 bp on a NextSeq 550 high-capacity chip.
376
377 Differential expression analysis
378 Demultiplexed reads were quality trimmed with Trimmomatic (v2.8.3) (Bolger et al. 2014),
379 duplicate reads with Piccard (Li et al. 2009), and mapped to the unmasked genome with
380 STAR(v2.7) (Dobin et al. 2013). Differential read analysis between conditions was assessed
381 with SARTools (v1.7.3) and DESeq2 (v1.12) (Varet et al. 2016). Lists of identified
382 differentially expressed genes were filtered with a Log_2 fold cut off of <-1.4 and >1.4 with an
383 FDR P_{adj} cut off of <0.05. Differential gene expression patterns between conditions was
384 assessed using GProfiler (e99_eg46_p14_f929183), DiVenn (v2.0), and STRING (v11.0)
385 (Pomaznoy et al. 2018; Sun et al. 2019; Szklarczyk et al. 2019).

386

387 SNP analysis

388 Trimmed RNASeq data used in differential expression analysis was mapped to the genome
389 using the GATK best practices pipeline that uses STAR(v2.7), Piccard (v2.22.2), Samtools
390 (v1.1), and GATK4 to generate a SNP VCF files for the HA and LA individuals (Li et al. 2009;
391 McKenna et al. 2010; Dobin et al. 2013; Brouard et al. 2019; GATK-Team 2019). Hard quality
392 filtering was performed on the VCF files to remove low coverage and poor-quality data
393 (DP<10, QD<5, MQ<40, SOR>3). Using SNPhylo (v 20160204) (maximum percent of sample
394 without SNP information (PNSS) 25 and max missing rate 0.5), a phylogenetic tree was
395 calculated, and visualised in FigTree (v1.4.3)(Lee et al. 2014; Rambaut 2016). Principal
396 component analysis (PCA) and Multi-dimensional scaling (MDS) were calculated with TASSEL
397 (v5.2.6) (Bradbury et al. 2007).

398

399 General Linear Model analysis was run on all detected SNPs, and the P-values plotted as a
400 Manhattan plot with TASSEL. F_{st} per SNP (MAC 8) between HA and LA populations was
401 calculated with VCFtools (v0.1.16). Tajima's Pi (per SNP) and Tajima's D (sliding window
402 30,000) were calculated for both HA and LA. Average F_{st} and -Log10 ratio of HA/LA Pi values
403 per gene were calculated and genes were subsequently filtered to retain the genes with the
404 top 10% of values in both. These were further filtered to include genes where HA-Pi was less
405 than or equal to half LA-Pi. Genes that had fewer than 5 SNPs were excluded. Genes that
406 passed these filters were assessed with STRING to identify connections to differentially
407 expressed genes identified in the differential gene expression analysis. Disconnected genes
408 were hidden and a high stringency filter used to mask low confidence gene-gene
409 interactions. To identify putative transcription factors and upstream pathway interactions of

genes identified in the STRING analysis, Transcription factor enrichment analysis (TFEA) and expression to kinases analysis was performed with eXpression2Kinases (Clarke et al. 2018). Putative genes were identified in differentially expressed genes and gProfiler gene enrichment performed.

Data availability

PRJNA623866 – Genome assembly

PRJNA638118 – Low altitude population transcriptome

PRJNA638117 - High altitude population transcriptome

Acknowledgements

Dr Stephen Short was supported by Natural Environment Research Council award number NE/R016429/1. Gratitude to Dr Paulino Sapo and the Serviço de Ambiente da Ilha do Pico or permission to research earthworms on the island.

Author contributions

I Perry: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualisation, Writing – original draft and Writing – review & editing.

SB Hernadi: Investigation and Methodology.

L Cunha: Investigation and Methodology.

S Short: Methodology and Writing – original draft.

A Marchbank: Resources.

D Spurgeon: Writing – original draft.

P Orozco-terWengel: Methodology, Supervision and Writing – original draft.

P Kille: Conceptualization, Data curation, Investigation, Methodology, Project administration, Supervision, Writing – original draft and Writing – review & editing.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

References and Citations

Ashe, T. 1813. *History of the Azores, or Western Islands*. London: Sherwood, Neely, and Jones, Paternoster Row.

Bart, S., Amosse, J., Lowe, C. N., Mougin, C., Pery, A. R. R. and Pelosi, C. 2018. Aporrectodea caliginosa, a relevant earthworm species for a posteriori pesticide risk assessment: current knowledge and recommendations for culture and experimental design. *Environ Sci Pollut Res Int* 25(34), pp. 33867-33881. doi: 10.1007/s11356-018-2579-9

Bhambri, A., Dhaunta, N., Patel, S. S., Hardikar, M., Srikakulam, N., Sivasubbu, S. and Pillai, B. 2017. Insights into regeneration from the genome, transcriptome and metagenome analysis of Eisenia fetida. *BioRxiv* 180612(1), doi: 10.1101/180612

453 Bianchi, M. E. 2009. HMGB1 loves company. *J Leukoc Biol* 86(3), pp. 573-576. doi:
454 10.1189/jlb.1008585
455
456 Blouin, M., Hodson, M. E., Delgado, E. A., Baker, G., Brussaard, L., Butt, K. R., Dai, J.,
457 Dendooven, L., Peres, G., Tondoh, J. E., et al. 2013. A review of earthworm impact on soil
458 function and ecosystem services. *European Journal of Soil Science* 64(2), pp. 161-182. doi:
459 10.1111/ejss.12025
460
461 Bolger, A. M., Lohse, M. and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina
462 Sequence Data. *Bioinformatics* 30(15), pp. 2114-2120.
463
464 Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. and Buckler, E. S.
465 2007. TASSEL: software for association mapping of complex traits in diverse samples.
466 *Bioinformatics* 23(19), pp. 2633-2635. doi: 10.1093/bioinformatics/btm308
467
468 Brouard, J. S., Schenkel, F., Marete, A. and Bissonnette, N. 2019. The GATK joint genotyping
469 workflow is appropriate for calling variants in RNA-seq experiments.*J Anim Sci Biotechnol*.
470 Vol. 10.
471
472 Butt, K. R. 2010. The earthworm inoculation unit technique: Development and use in soil
473 improvement over two decades. *Biology of Earthworms* 24(1),
474
475 Butt, K. R. and Briones, M. J. I. 2011. Life cycle studies of the earthworm Lumbricus friendi
476 (Cognetti, 1904). *Pedobiologia* 54(1), pp. S27-S29. doi: 10.1016/j.pedobi.2011.08.003

Carine, M. A. and Schaefer, H. 2009. The Azores diversity enigma: why are there so few Azorean endemic flowering plants and why are they so widespread? *Journal of Biogeography* 37(1), pp. 77-89. doi: 10.1111/j.1365-2699.2009.02181.x

Chang, C. H. and James, S. 2011. A critique of earthworm molecular phylogenetics. *Pedobiologia* 54, pp. S3-S9. doi: 10.1016/j.pedobi.2011.07.015

Clarke, D. J., Kuleshov, M. V., Schilder, B. M., Torre, D., Duffy, M. E., Keenan, A. B., Lachmann, A., Feldmann, A. S., Gundersen, G. W., Silverstein, M. C., et al. 2018. eXpression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks.*Nucleic Acids Res*. Vol. 46. pp. W171-179.

Concetti, J. and Wilson, C. L. 2018. NFKB1 and Cancer: Friend or Foe?*Cells*. Vol. 7.

Costanzo, J. P. and Lee, R. E., Jr. 2013. Avoidance and tolerance of freezing in ectothermic vertebrates. *J Exp Biol* 216(Pt 11), pp. 1961-1967. doi: 10.1242/jeb.070268

Cunha, L., Campos, I., Montiel, R., Rodrigues, A. and Morgan, A. J. 2011. Morphometry of the epidermis of an invasive megascoelecid earthworm (Amynthas gracilis, Kinberg 1867) inhabiting actively volcanic soils in the Azores archipelago. *Ecotoxicol Environ Saf* 74(1), pp. 25-32. doi: 10.1016/j.ecoenv.2010.08.004

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1), pp. 15-21. doi: 10.1093/bioinformatics/bts635

Drummond, A. J., Suchard, M. A., Xie, D. and Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8), pp. 1969-1973. doi: 10.1093/molbev/mss075

Eloy de Amorim, M., Schoener, T. W., Santoro, G., Lins, A. C. R., Piovia-Scott, J. and Brandao, R. A. 2017. Lizards on newly created islands independently and rapidly adapt in morphology and diet. *Proc Natl Acad Sci U S A* 114(33), pp. 8812-8816. doi: 10.1073/pnas.1709080114

Gale, J. 2004. Plants and altitude--revisited. *Ann Bot* 94(2), p. 199. doi: 10.1093/aob/mch143

GATK-Team. 2019. *GATK - RNAseq short variant discovery (SNPs + Indels)*. https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels-: Broad Instutute. Available at: http://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels- [Accessed: 29/02/20].

Ghurye, J. and Pop, M. 2019. Modern technologies and algorithms for scaffolding assembled genomes. *PLoS Comput Biol* 15(6), p. e1006994. doi: 10.1371/journal.pcbi.1006994

Giardina, B., Chiancone, E. and Antonini, E. 1975. Studies on erythrocruorin. *Journal of Molecular Biology* 93(1), pp. 1-10. doi: 10.1016/0022-2836(75)90355-1

Gonzalez, G., Garcia, E., Cruz, V., Borges, S., Zalamea, M. and Rivera, M. M. 2007. Earthworm communities along an elevation gradient in Northeastern Puerto Rico. *European Journal of Soil Biology* 43, pp. S24-S32. doi: 10.1016/j.ejsobi.2007.08.044

Jin, F., Zhou, Z., Guo, Q., Liang, Z., Yang, R., Jiang, J., He, Y., Zhao, Q. and Zhao, Q. 2020. High-quality genome assembly of Metaphire vulgaris. *PeerJ* 8, p. e10313. doi: 10.7717/peerj.10313

Kanchilakshmi, M. 2016. Earthworm: A Potential and Sustainable Source for Soil Fertility – An Altitude Based Biophysical Study. *International Journal of Environmental Protection and Policy* 4(3), p. 77. doi: 10.11648/j.ijepp.20160403.15

Kang, R., Chen, R., Zhang, Q., Hou, W., Wu, S., Cao, L., Huang, J., Yu, Y., Fan, X. G., Yan, Z., et al. 2014. HMGB1 in health and disease. *Mol Aspects Med* 40, pp. 1-116. doi: 10.1016/j.mam.2014.05.001

Ke, Q. and Costa, M. 2006. Hypoxia-inducible factor-1 (HIF-1). *Mol Pharmacol* 70(5), pp. 1469-1480. doi: 10.1124/mol.106.027029

Kumar, S., Stecher, G. and Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Molecular Biology and Evolution* 33(7), pp. 1870-1874.

Laetsch, D. R. and Blaxter, M. L. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research 2017 6:1287* 6(1287), doi: doi:10.12688/f1000research.12232.1

Lee, T. H., Guo, H., Wang, X., Kim, C. and Paterson, A. H. 2014. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* 15(1),

Lenoir, J., Gegout, J. C., Marquet, P. A., de Ruffray, P. and Brisse, H. 2008. A significant upward shift in plant species optimum elevation during the 20th century. *Science* 320(5884), pp. 1768-1771. doi: 10.1126/science.1156831

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16), pp. 2078-2079. doi: 10.1093/bioinformatics/btp352

Lyman, R. F., Nevo, E. and Mackay, T. F. 2002. Variation in Drosophila sensory bristle number at 'Evolution Canyon'. *Genet Res* 80(3), pp. 215-223. doi: 10.1017/s0016672302005876

Majmundar, A. J., Wong, W. J. and Simon, M. C. 2010. Hypoxia-inducible factors and the response to hypoxic stress. *Mol Cell* 40(2), pp. 294-309. doi: 10.1016/j.molcel.2010.09.022

Manalo, D. J., Rowan, A., Lavoie, T., Natarajan, L., Kelly, B. D., Ye, S. Q., Garcia, J. G. and Semenza, G. L. 2005. Transcription regulation of vascular endothelial cell responses to hypoxia by HIF-1. *Blood* 105(2), pp. 659-669.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M. A. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9), pp. 1297-1303. doi: 10.1101/gr.107524.110

Millien, V. 2006. Morphological evolution is accelerated among island mammals. *PLoS Biol* 4(10), p. e321. doi: 10.1371/journal.pbio.0040321

Miranda, J. M., Mendes-Vicotor, L. A., Simoes, J. Z., Luis, J. F., Matias, L., Shimamura, H., Shiobara, H., Nemoto, H., Mochizuki, H., Hirn, A., et al. 1998. Tectonic setting of the Azores Plateau deduced from a OBS survey. *Marine Geophysical Researches* 20(3), pp. 171-182. doi: Doi 10.1023/A:1004622825210

Muller, S., Ronfani, L. and Bianchi, M. E. 2004. Regulated expression and subcellular localization of HMGB1, a chromatin protein with a cytokine function. *J Intern Med* 255(3), pp. 332-343. doi: 10.1111/j.1365-2796.2003.01296.x

Novo, M., Cunha, L., Maceda-Veiga, .A, Talavera, J. A., Hodson, M. E., Spurgeon, D., Bruford, M. W., Morgan, A. J. and Kille, P. 2015. Multiple introductions and environmental factors affecting the establishment of invasive species on a volcanic island. *Soil Biology & Biochemistry* 85(1), pp. 89-100. doi: 10.1016/j.soilbio.2015.02.031

Pacheco, P. R., Branco, C. C., Gomes, C. T., Cabral, R. and Mota-Vieira, L. 2010. HLA Class I and II profiles in Sao Miguel Island (Azores): genetic diversity and linkage disequilibrium. *BMC Res Notes* 3(1), p. 134. doi: 10.1186/1756-0500-3-134

Peacock, A. J. 1998. ABC of oxygen: oxygen at high altitude. *BMJ* 317(7165), pp. 1063-1066. doi: 10.1136/bmj.317.7165.1063

Pomaznoy, M., Ha, B. and Peters, B. 2018. GOnet: a tool for interactive Gene Ontology analysis. *BMC Bioinformatics* 19(470),

Powell, J. P. and Reinhard, S. 2016. Measuring the effects of extreme weather events on yields. *Weather and Climate Extremes* 12(1), pp. 69-79. doi: 10.1016/j.wace.2016.02.003

Pérez-Losada, M., Ricoy, M., Marshall, J. C. and Domínguez, J. 2009. Phylogenetic assessment of the earthworm Aporrectodea caliginosa species complex (Oligochaeta: Lumbricidae) based on mitochondrial and nuclear DNA sequences. *Mol Phylogenet Evol* 52(2), pp. 293-302. doi: 10.1016/j.ympev.2009.04.003

Rambaut, A. 2016. *FigTree v1.4.3.* http://tree.bio.ed.ac.uk/software/figtree/: Available at: [Accessed: 15/01/18].

Rimington, O. J. 2022. *The mechanisms of evolutionary flexibility in earthworm genomes.* Cardiff University.

Rimmington, O. 2019. Nanochrome.
https://github.com/OliverCardiff/Nanochrome_scaffolder.

Robledo, D., Rubiolo, J. A., Cabaleiro, S., Martinez, P. and Bouza, C. 2017. Differential gene expression and SNP association between fast- and slow-growing turbot (Scophthalmus maximus). *Sci Rep* 7(1), p. 12105. doi: 10.1038/s41598-017-12459-4

Rodrigues, A. F., Martins, N. O., Ribeiro, N. and Joaquinito, A. 2015. Early Atlantic navigation: pre-Portuguese p resence in the Azores islands. *Archaeological Discovery* 3(1), pp. 104-113.

Royer Jr, W. E., Sharma, H., Strand, K., Knapp, J. E. and Bhyravbhatla, B. 2006. Lumbricus Erythrocruorin at 3.5 Å Resolution: Architecture of a Megadalton Respiratory Complex. *Structure* 14(7), pp. 1167-1177.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E. and Sanchez-Gracia, A. 2017. DnaSP v6: DNA sequence polymorphism analysis of large datasets. *Molecular Biology and Evolution* 34(1), pp. 3299-3302.

Ruan, J. and Li, H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* 17(2), pp. 155-158. doi: 10.1038/s41592-019-0669-3

Scheinfeldt, L. B. and Tishkoff, S. A. 2010. Living the high life: high-altitude adaptation. *Genome Biol* 11(9), p. 133. doi: 10.1186/gb-2010-11-9-133

642 Schubart, C. D., Diesel, R. and Hedges, S. B. 1998. Rapid evolution to terrestrial life in
643 Jamaican crabs. *Nature* 393(6683), pp. 363-365. doi: 10.1038/30724
644
645 Shao, Y., Wang, X-B., Zhang, J-J., Li, M-L., Wu, S-S., Ma, X-Y., Wang, X., Zhao, H-F., Li, Y.,
646 He, H. Z., et al. 2020. Genome and single-cell RNA-sequencing of the earthworm Eisenia
647 andrei identifies cellular mechanisms underlying regeneration. *Nat Commun* 11(1), p. 2656.
648 doi: 10.1038/s41467-020-16454-8
649
650 Singh, J., Schadler, M., Demetrio, W., Brown, G. G. and Eisenhauer, N. 2019. Climate change
651 effects on earthworms - a review. *Soil org.* 91(3), pp. 114-138.
652
653 Skarbø, K. and VanderMolen, K. 2015. Maize migration: key crop expands to higher altitudes
654 under climate change in the Andes. *Climate and Development* 8(3), pp. 245-255. doi:
655 10.1080/17565529.2015.1034234
656
657 Smit, A. F. A., Hubley, R. and Green, P. 2015. RepeatMasker Open-4.0 2013-2015.
658 www.repeatmasker.org: ISB.
659
660 Somero, G. N. 2005. Linking biogeography to physiology: Evolutionary and acclimatory
661 adjustments of thermal limits. *Front Zool* 2(1), p. 1. doi: 10.1186/1742-9994-2-1
662
663 Storz, J. F. 2016. Hemoglobin-oxygen affinity in high-altitude vertebrates: is there evidence
664 for an adaptive trend? *Journal of Experimental Biology* 219, pp. 3190-3203.
665

666 Storz, J. F., Scott, G. R. and Cheviron, Z. A. 2010. Phenotypic plasticity and genetic
667 adaptation to high-altitude hypoxia in vertebrates. *J Exp Biol* 213(Pt 24), pp. 4125-4136. doi:
668 10.1242/jeb.048181
669
670 Sun, L., Dong, S., Ge, Y., Fonseca, J. P., Robinson, Z., Mysore, K. and Mehta, P. 2019. DiVenn:
671 An interactive and integrated Web-based visualization tool for comparing gene lists.
672 *Frontiers in Genetics* 10(421),
673
674 Szklarczyk, D. et al. 2019. STRING v11: protein-protein association networks with increased
675 coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic*
676 *Acids Research* 47(D607-613),
677
678 Tolessa, K., D'Heer, J., Duchateau, L. and Boeckx, P. 2017. Influence of growing altitude,
679 shade and harvest period on quality and biochemical composition of Ethiopian specialty
680 coffee. *J Sci Food Agric* 97(9), pp. 2849-2857. doi: 10.1002/jsfa.8114
681
682 Varet, H., Brillet-Gueguen, L., Coppee, J. Y. and Dillies, M. A. 2016. SARTools: A DESeq2- and
683 EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS One*
684 11(6), p. e0157022. doi: 10.1371/journal.pone.0157022
685
686 Wang, J. R., Holt, J., McMillan, L. and Jones, C. D. 2018. FMLRC: Hybrid long read error
687 correction using an FM-index. *BMC Bioinformatics* 19(1), p. 50. doi: ARTN 50
688 10.1186/s12859-018-2051-3
689

690 Wang, X., Zhang, Y., Zhang, Y., Kang, M., Li, Y., James, S. W., Yang, Y., Bi, Y., Jiang, H., Zhao, Y.
691 and Sun, Z. 2021. Amynthas corticis genome reveals molecular mechanisms behind global
692 distribution. *Commun Biol* 4(1), p. 135. doi: 10.1038/s42003-021-01659-4
693
694 Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G.,
695 Kriventseva, E. V. and Zdobnov, E. M. 2017. BUSCO applications from quality assessments to
696 gene prediction and phylogenomics. *Molecular Biology and Evolution* 35(3), pp. 543-548.
697
698 Woodhall, D. 1974. Geology and Volcanic History of Pico Island Volcano, Azores. *Nature*
699 248(5450), pp. 663-665. doi: DOI 10.1038/248663a0
700
701 Xu, G.-C., Xu, T.-J., Zhu, R., Zhang, Y., Li, S.-Q., Wang, H.-W. and Li, J.-T. 2018. LR-Gapcloser: a
702 tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience*
703 8, pp. 1-14.
704
705 Xue, W., Li, J. T., Zhu, Y. P., Hou, G. Y., Kong, X. F., Kuang, Y. Y. and Sun, X. W. 2013.
706 L_RNA_scaffolder: scaffolding genomes with transcripts. *BMC Genomics* 14, p. 604. doi:
707 10.1186/1471-2164-14-604
708
709 Zorn, M. I., Van Gestel, C. A. M. and Eijsackers, H. 2005. Species-specific earthworm
710 population responses in relation to flooding dynamics in a Dutch floodplain soil.
711 *Pedobiologia* 49(3), pp. 189-198. doi: 10.1016/j.pedobi.2004.08.004
712

713 Zwarycz, A. S., Nossa, C. W., Putnam, N. H. and Ryan, J. F. 2015. Timing and Scope of
714 Genomic Expansion within Annelida: Evidence from Homeoboxes in the Genome of the
715 Earthworm Eisenia fetida. *Genome Biol Evol* 8(1), pp. 271-281. doi: 10.1093/gbe/evv243

716

717 **Figure Legends**

718 **Figure 1: Population geography and history.**

719 **(A)** Approximate location and altitude of dig sites for island and high altitude (HA)
720 experimental and Native expression individuals (α), low altitude (LA) experimental
721 individuals (β) and LA native expression individuals (γ). **(B)** Long-term population decline of
722 *A. caliginosa* started during the last glacial maximum. Bayesian skyline plot analysis of COII
723 sequences drawn from individuals from across the island (Table S1) show a Median value
724 and range between upper and lower confidence interval shaded in blue. **(C)** Dated
725 phylogeny estimating divergence time for *A. caliginosa*, calculated using COII haplotype
726 groups with trees generated on BEAST v1.8.4. and visualised with TreeAnnotator v2.4.8.
727 Time scale is in 6 million years and blue node bars indicate the confidence interval of branch
728 splits.

729 **Figure 2: Genome assembly and annotation statistics.**

730 **(A)** Assembly statistics of the assembled *A. caliginosa* genome. Genome size is 1.1 Gbp,
731 longest contig is 6.5 Mbp and N50 is 1.2 Mbp. 279 contigs account for 50%f of the genome
732 assembly. 530 contigs account for 90% of the genome. GC composition is 40.2%. **(B)**
733 Identification of bacterial contamination including verminephrobacter (red circle) with
734 Blobtools prior to removal from genome assembly. **(C)** Identification of genome annotations

and the number of genes with multiple occurrences within the genome reveals most of the identified genes occur only once.

Figure 3: Differential gene expression analysis.

(A) Principal component analysis plot of all experimental and native individuals. High Altitude (HA) individuals (blue circle) separate cleanly from the Low altitude (LA) individuals (red circle) via PC3. B,C,D) Identification of shared differentially regulated genes for both HA and LA individuals (upregulate red, down regulated blue, conflicting yellow). Conditions investigated: **(B)** Temperature at three Oxygen concentrations, **(C)** Oxygen concentrations at 4°C and **(D)** Oxygen concentrations at 21°C.

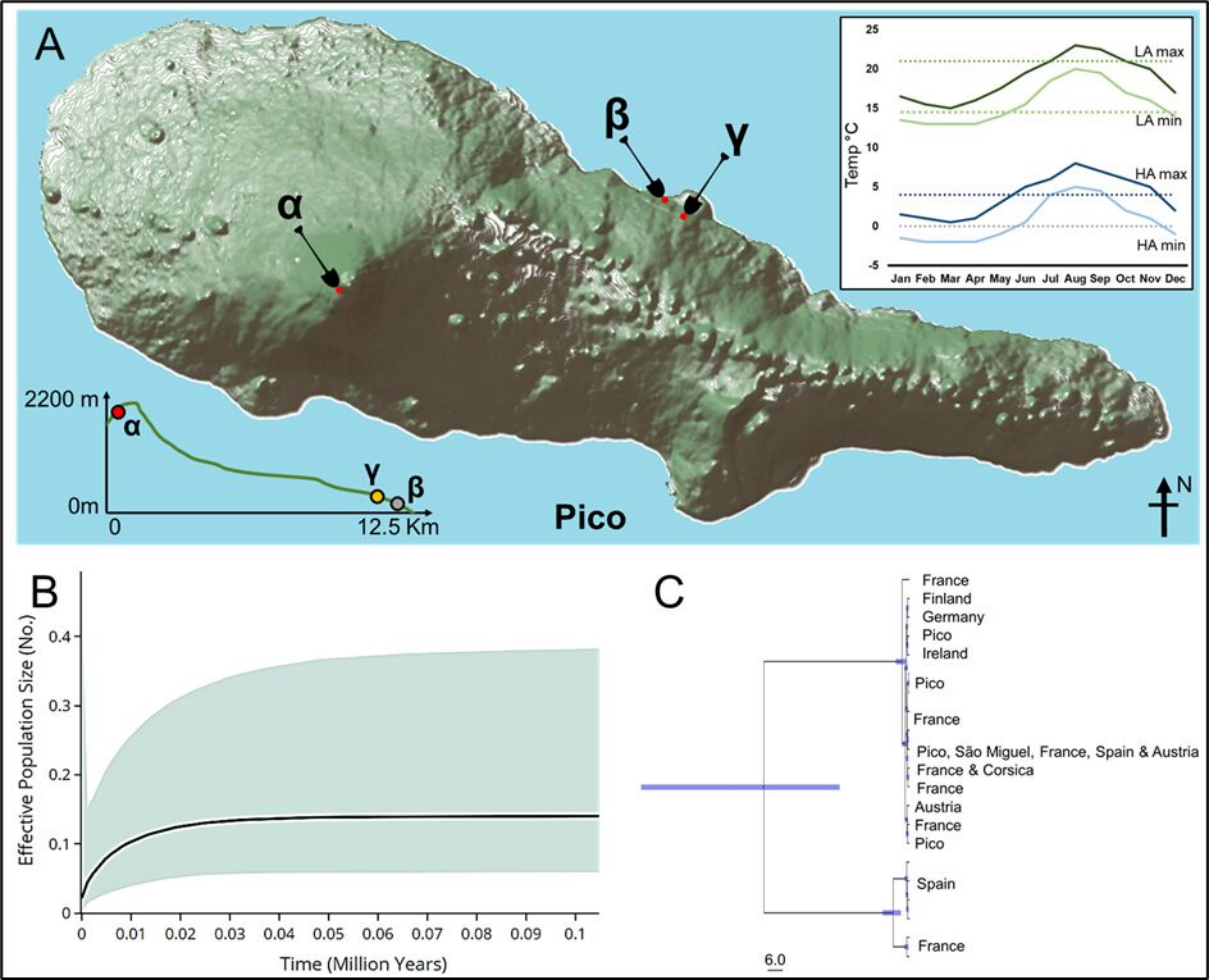
Figure 4: SNP analysis from transcriptomic reads.

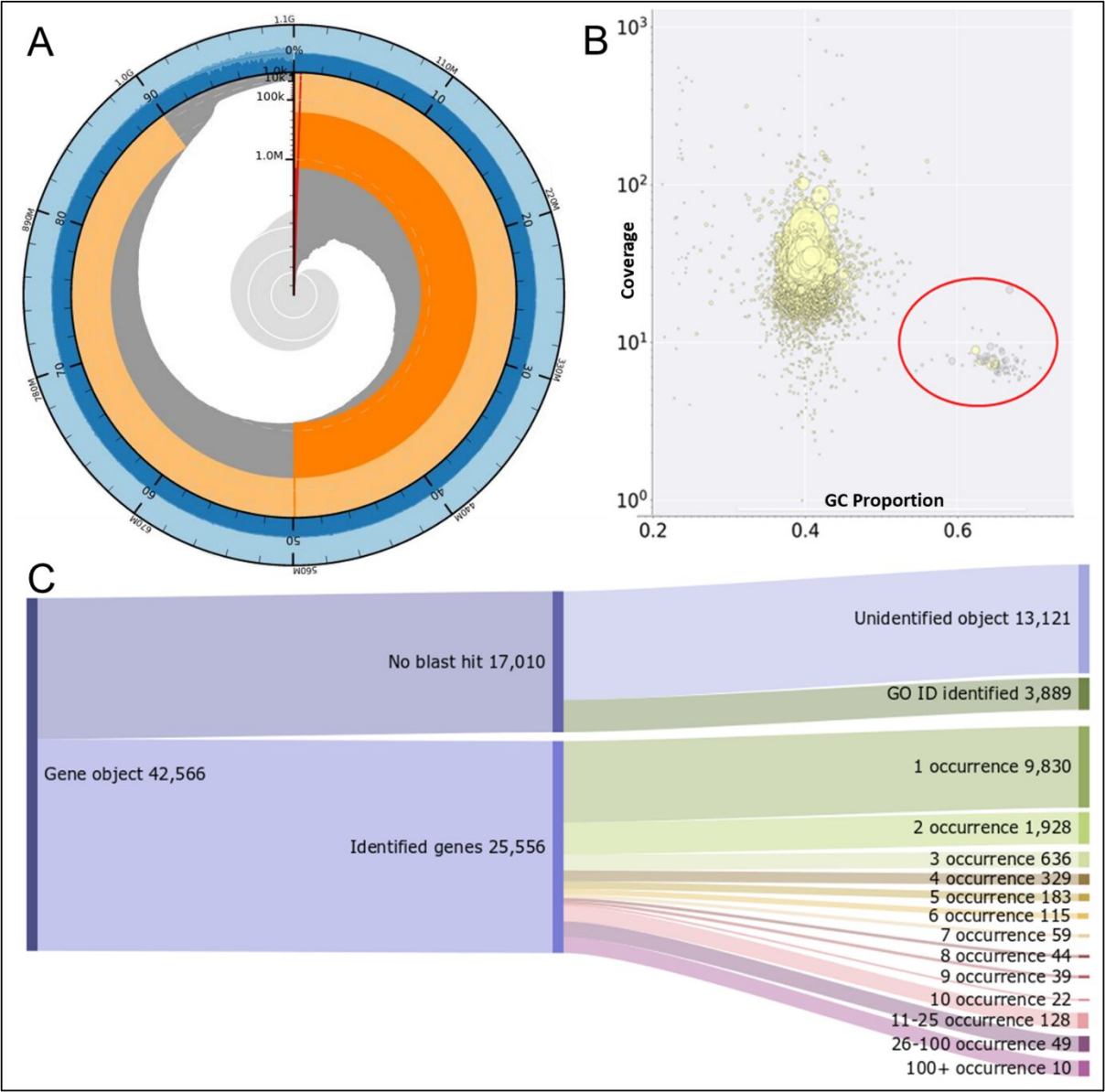
(A) SNP phylogeny separation of High altitude (HA) individuals (experimental red, native orange) from low altitude (LA) individuals (experimental blue, native cyan). LA natives taken from a separate low altitude site are slightly diverged from LA experimental individuals. **(B)** PCA separation of HA and LA individuals. **(C)** MDS separation of HA and LA individuals. **(D)** SNP P-value distribution across the genome as a moving average indicating areas of elevated diversity. **(E)** Filtering distribution of SNPs with Average FST per gene and Average nucleotide diversity per gene.

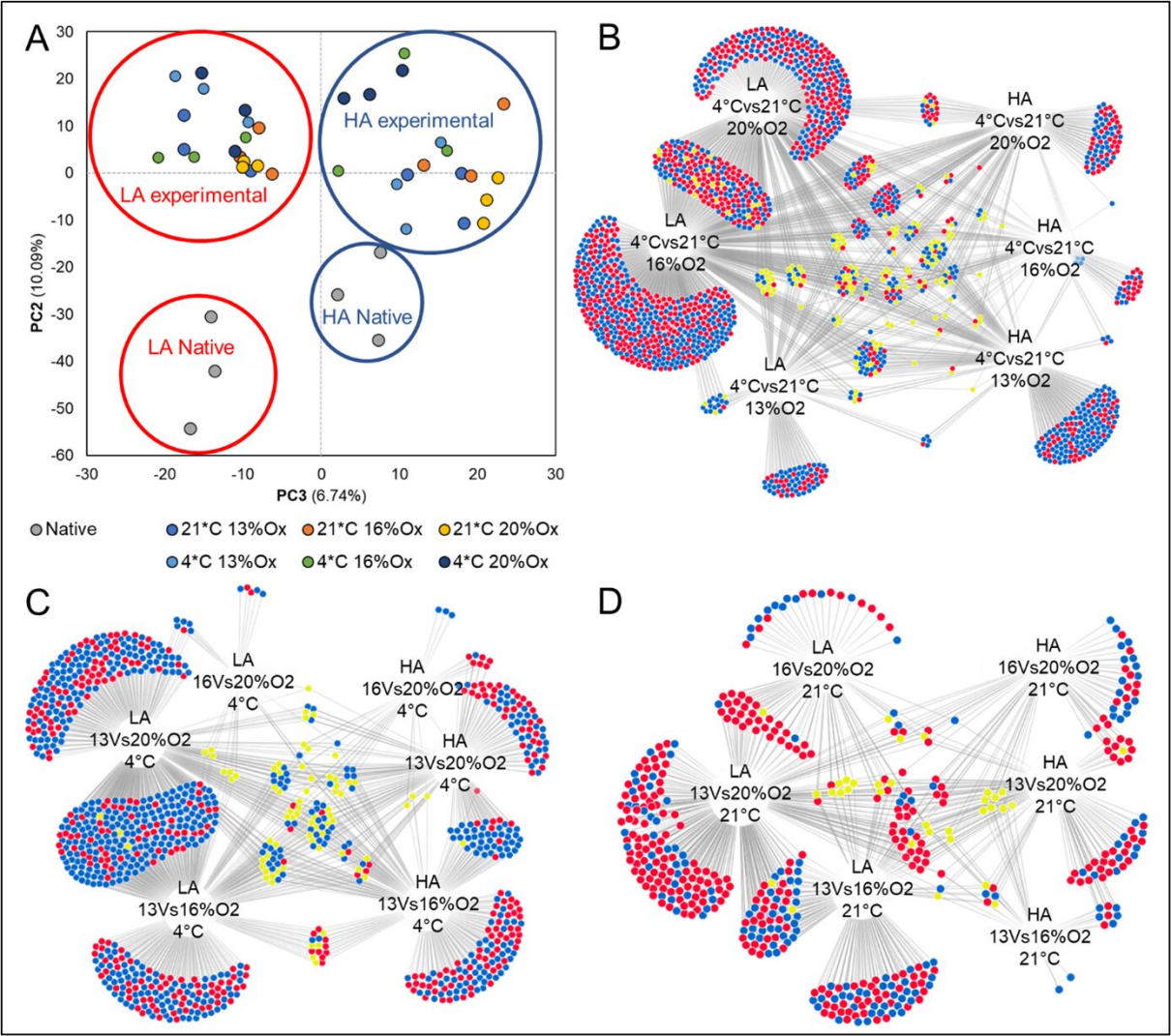
Figure 5: Genes with high number of SNPs have direct interactions with the differentially regulated genes.

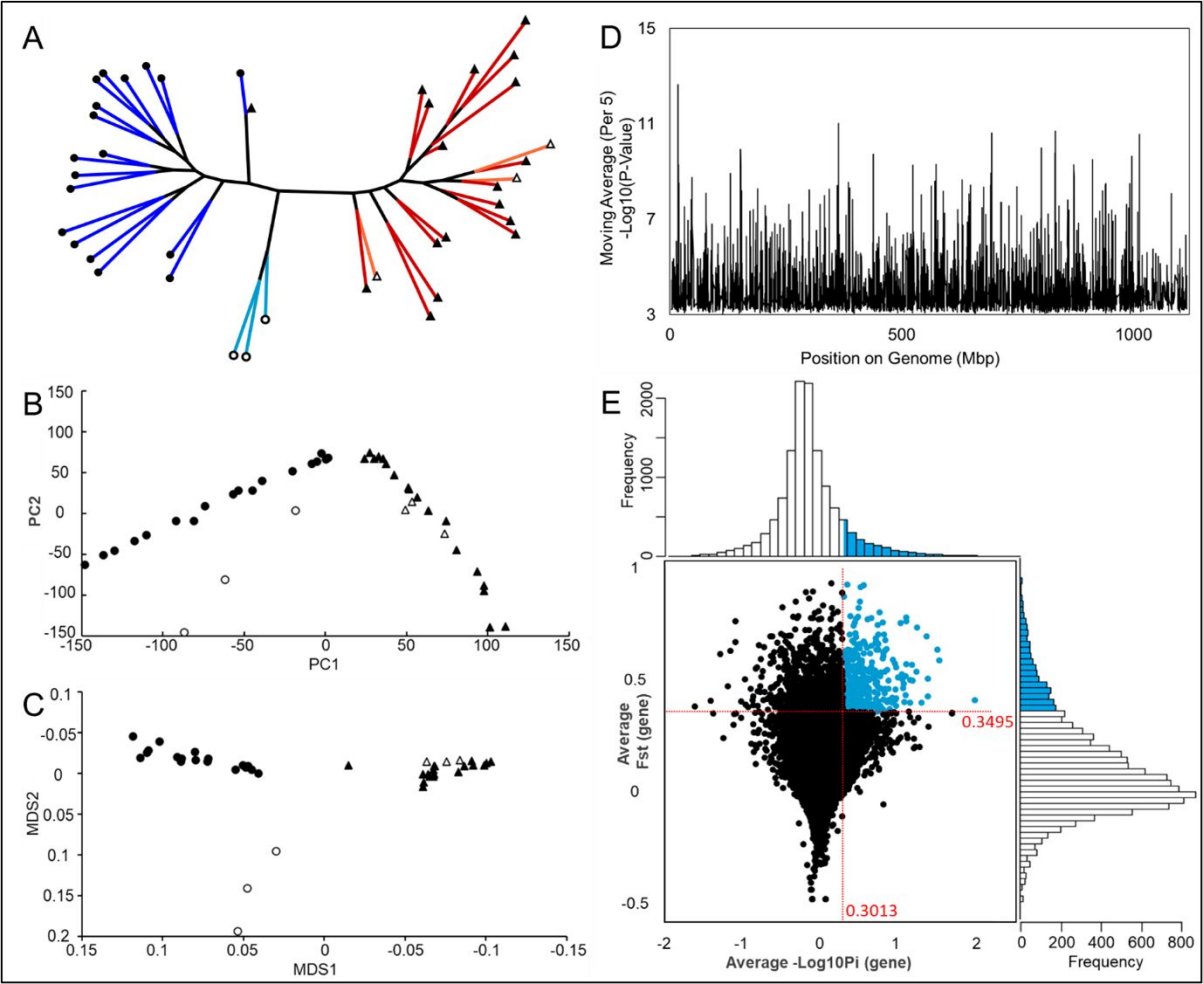
(A) Kmeans clustering of HMGB1, HMGB2 (pink) and TP53 (cyan) interacting with genes under selective pressure determined by STRING-DB (STRING genes) with high confidence (0.7) minimum interaction score. HMB1, HMGB2, TP53 and cluster 1 genes (green) are

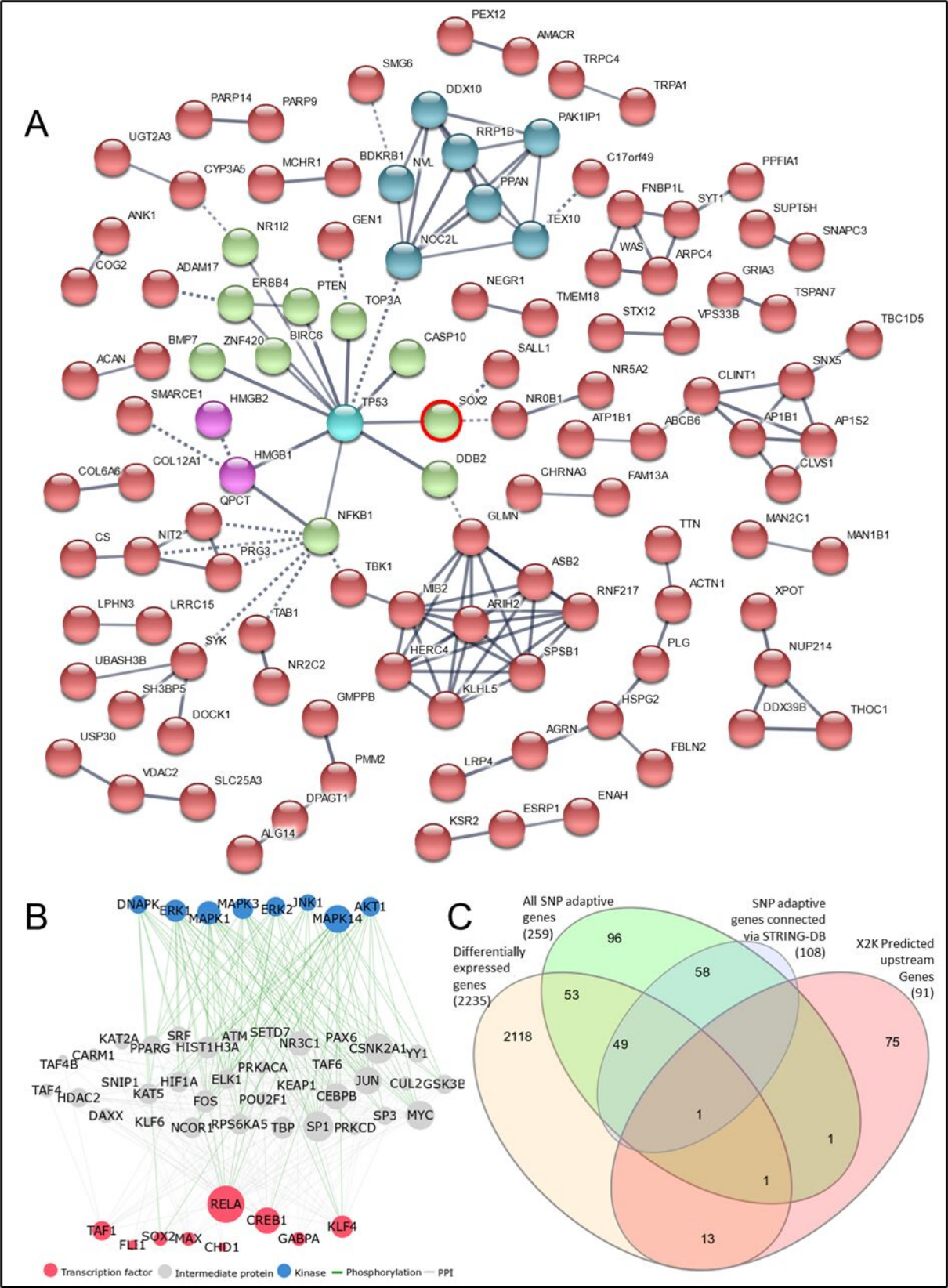
responsible for ‘DNA topological change’, cluster 2 genes (blue) are responsible for ‘Ribosome biogenesis’, and TBK1, GLMN, MB2, ASB2, ARIH2, RNF217, SPSB1, HERC4 and KLHL5 within cluster 3 (red) are responsible for ‘Protein ubiquitination’. **(B)** X2K interaction of kinases, TFs and upstream activation pathway to STRING genes. **(C)** Overlaps between Differentially expressed genes, genes under selective pressure, STRING genes and genes upstream of STRING identified by X2K.


A
α
β
γ
LA max
LA min
HA max
HA min
Temp °C
25
20
15
10
5
0
-5
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2200 m
0m
0
12.5 Km
Pico
N
B
Effective Population Size (No.)
0.4
0.3
0.2
0.1
0
0 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.1
Time (Million Years)
C
France
Finland
Germany
Pico
Ireland
Pico
France
Pico, São Miguel, France, Spain & Austria
France & Corsica
France
Austria
France
Pico
Spain
France
6.0


A
B
C
Coverage
GC Proportion
No blast hit 17,010
Gene object 42,566
Identified genes 25,556
Unidentified object 13,121
GO ID identified 3,889
1 occurrence 9,830
2 occurrence 1,928
3 occurrence 636
4 occurrence 329
5 occurrence 183
6 occurrence 115
7 occurrence 59
8 occurrence 44
9 occurrence 39
10 occurrence 22
11-25 occurrence 128
26-100 occurrence 49
100+ occurrence 10


A
LA experimental
HA experimental
HA Native
LA Native
PC2 (10.09%)
PC3 (6.74%)
Native
21*C 13%Ox
21*C 16%Ox
21*C 20%Ox
4*C 13%Ox
4*C 16%Ox
4*C 20%Ox
B
LA 4°Cvs21°C 20%O2
HA 4°Cvs21°C 20%O2
LA 4°Cvs21°C 16%O2
HA 4°Cvs21°C 16%O2
LA 4°Cvs21°C 13%O2
HA 4°Cvs21°C 13%O2
C
LA 16Vs20%O2 4°C
HA 16Vs20%O2 4°C
LA 13Vs20%O2 4°C
HA 13Vs20%O2 4°C
LA 13Vs16%O2 4°C
HA 13Vs16%O2 4°C
D
LA 16Vs20%O2 21°C
HA 16Vs20%O2 21°C
LA 13Vs20%O2 21°C
HA 13Vs20%O2 21°C
LA 13Vs16%O2 21°C
HA 13Vs16%O2 21°C


A
B
PC2
PC1
C
MDS2
MDS1
D
Moving Average (Per 5)
-Log10(P-Value)
Position on Genome (Mbp)
E
Frequency
Average
Fst (gene)
0.3495
0.3013
Average -Log10Pi (gene)
Frequency


A
PEX12
AMACR
SMG6
TRPC4
TRPA1
PARP14
PARP9
DDX10
PAK1IP1
RRP1B
UGT2A3
CYP3A5
MCHR1
BDKRB1
NVL
PPAN
C17orf49
PPFIA1
FNBP1L
SYT1
ANK1
NR1I2
GEN1
TEX10
SUPT5H
SNAPC3
NOC2L
WAS
ARPC4
COG2
ADAM17
ERBB4
PTEN
TOP3A
NEGR1
GRIA3
TMEM18
TSPAN7
STX12
VPS33B
TBC1D5
BMP7
ZNF420
BIRC6
CASP10
SALL1
ACAN
SNX5
NR5A2
CLINT1
SMARCE1
HMGB2
TP53
SOX2
NR0B1
AP1B1
AP1S2
ATP1B1
ABCB6
COL12A1
HMGB1
DDB2
CLVS1
COL6A6
CHRNA3
FAM13A
QPCT
NFKB1
GLMN
TTN
MAN2C1
CS
NIT2
PRG3
MAN1B1
TBK1
ASB2
ACTN1
MIB2
RNF217
XPOT
LPHN3
LRRC15
TAB1
ARIH2
PLG
SYK
NUP214
UBASH3B
HERC4
SPSB1
NR2C2
SH3BP5
KLHL5
HSPG2
DDX39B
THOC1
GMPPB
DOCK1
AGRN
USP30
FBLN2
PMM2
LRP4
VDAC2
SLC25A3
DPAGT1
ENAH
ESRP1
ALG14
KSR2
B
DNAPK
ERK1
MAPK3
ERK2
JNK1
AKT1
MAPK1
MAPK14
KAT2A
SRF
ATM
SETD7
PAX6
PPARG
HIST1H3A
NR3C1
CSNK2A1
YY1
CARM1
TAF4B
PRKACA
TAF6
SNIP1
HIF1A
ELK1
KEAP1
JUN
CUL2
GSK3B
TAF4
HDAC2
KAT5
FOS
POU2F1
CEBPB
DAXX
KLF6
NCOR1
RPS6KA5
TBP
SP1
PRKCD
SP3
MYC
RELA
CREB1
KLF4
TAF1
FLI1
SOX2
MAX
CHD1
GABPA
Transcription factor
Intermediate protein
Kinase
Phosphorylation
PPI
C
All SNP adaptive genes (259)
SNP adaptive genes connected via STRING-DB (108)
X2K Predicted upstream Genes (91)
Differentially expressed genes (2235)
96
58
53
2118
49
75
1
1
1
13