Title

Role of aridity in shaping adaptive genomic divergence and population connectivity in a Southern African rodent

Abstract

Elucidating the drivers of evolution in dry environments is central to understanding how organisms respond to climate change. While research on the genomics of adaptation is growing, aridity-driven intraspecific divergence remains poorly quantified. Here, we address this gap by using genomic data from 230 individuals of the arid-adapted four-striped mouse *Rhabdomys bechuanae*, sampled across an aridity gradient in southern Africa, a region facing increasing aridification. Combining these data with palaeoclimatic reconstructions and present-day aridity indices, we investigate, from a spatio-temporal perspective, how intraspecific genetic variation relates to aridity. Inference of past effective population size revealed a sharp decline in the late Pleistocene, coinciding with regional aridification and potentially reflecting changes in connectivity during dry periods. Current population structure followed a pattern of isolation by distance and mirrored the aridity gradient. Genotype-Environment Association analyses identified SNPs and genes significantly associated with aridity and genetically differentiated among populations, with functions related to water and energy conservation - as expected under arid conditions - as well as neurotransmission. These findings highlight the underappreciated role for neurological processes in coping with water and resource scarcity. More broadly, our integrative genomics approach suggests that aridity shapes population connectivity and adaptation, with implications for climate resilience.

Keywords

adaptation, aridity, effective population size, genotype-environment association, population genomics, *Rhabdomys bechuanae*

Introduction

Climate change has increasingly been associated with local extinctions [1], shifts in phenology [2] and alteration of species' geographic distributions [3]. These responses tend to be particularly pronounced in harsh environments, such as arid regions, where species already face several ecological and physiological constraints [4]. With projections indicating a rise in the intensity and frequency of drought events in the coming decades [5,6], it is essential to better understand how species respond to both temporal and spatial variation in aridity. By combining newly generated genomic data with available palaeoclimatic reconstructions and contemporary measures of aridity, this study aims to shed light on the role of aridification on the demographic history, spatial population structure and adaptive responses of a non-model rodent species. These insights are key for predicting resilience of natural populations under future climate scenarios and for guiding conservation strategies [7].

At the individual level, the primary challenges of life in arid environments include maintaining body temperature and conserving water [8]. These challenges are especially evident for species that rely on evaporative water loss mechanisms, such as sweating, to dissipate heat, particularly in mammals [9]. The physiological demands imposed by such environmental conditions, particularly when experienced over multiple generations, often exceed the limits of phenotypic plasticity in non-desert-adapted species. Consequently, it is assumed that strong selection has acted on complex physiological traits related to energy metabolism, water retention and thermoregulation in species inhabiting arid environments [8-10].

Genomic studies have identified recurrent genetic patterns among mammals living in deserts [10]. Evidence of adaptation to arid environments has included elevated allele frequency differentiation [11,12], reduced genetic diversity [11,13-15], increased non-synonymous-to-synonymous substitution rate ratios [16] and divergent gene expression patterns [17] in genomic regions involved in key functions for desert adaptation. These genomic regions often relate to key physiological traits relevant to the survival of both mammalian specialist and non-specialist species in deserts, including DNA repair, protein synthesis and degradation [15,16], thyroid hormone levels [11], water retention [18], lipid [16] and carbohydrate metabolism [11,13,19-20]. However, most of these studies have relied on qualitative comparisons between desert and non-desert species or populations, and relatively few have explicitly explored quantitative variation along continuous environmental gradients. Several other studies on the genomics of local adaptation have demonstrated the power of Genotype-Environment Association (GEA) analyses to identify polymorphisms associated with continuous environmental variation [21-23].

However, continuous variation in precipitation or temperature does not, in of itself, fully capture the extent of the ecological pressures affecting homeotherms in an arid environment. Correlating allelic frequencies with temperature alone has the limitation of not taking into account precipitation, and vice versa; high temperatures in wetter areas do not affect organisms in the same way as they do in drier ones. These limitations are addressed by using a summary variable; in this context, Rocha et al. [10] highlighted the

use of the Aridity Index (AI), known to be a good predictor of certain aridity-adapted phenotypes, as a powerful, composite and quantitative environmental variable to inform genomic studies of adaptation to arid conditions. Aridity Index has been shown to be a major driver of ecosystem structure and functioning in drylands, and exhibits important ecological shifts along gradients of aridity [24].

In this study, we applied a GEA approach to investigate the genomic signatures of adaptation to arid conditions in *Rhabdomys bechuanae*, a four-striped mouse species from Southern Africa. Its range overlaps with hot arid and hyper-arid summer rainfall regions of Southern Africa [25]. Divergence from its closest relatives within the *Rhabdomys* species complex occurred approximately 3.1-4.3 million years ago [26], and this diversification has been associated with environmental niche differentiation [27]. It likely reflects a long-term trend towards increasing aridity and seasonality in Southern Africa since the late Miocene, coupled with a shift from closed subtropical woodland to sparse, shrubby vegetation [28]. Currently, *R. bechuanae* thrives predominantly in sparsely vegetated areas and nests in bushes. Phenotypic traits suggest adaptation to aridity; these traits include morphological (longer tail length related to potential differences in locomotion and/or thermoregulation) and behavioural divergence (selection of different habitats) from more mesic congeners [23,29], as well as enhanced osmoregulatory ability, as suggested by differences in blood metabolite concentrations between populations of *R. bechuanae* and of its sister species *R. dilectus,* occurring syntopically in semi-arid environments [30]. Although *R. bechuanae* is considered one of the most arid-adapted species of its genus, it still occupies a wide environmental gradient, providing an ideal system to investigate whether the aridity gradient imposes divergent selection pressures on key physiological traits. We therefore hypothesised that this aridity gradient would impose divergent selective pressures on key physiological traits related to adaptation to aridity, leading to detectable intraspecific population differentiation and genetic variation at loci associated with aridity-related functions.

To investigate genomic signatures of evolution under arid conditions across an extensive sample of individuals (n=230) from multiple localities spanning *R. bechuanae*'s distribution along an aridity gradient, we generated a genome-wide SNP dataset using restriction-site associated DNA sequencing (RAD-seq) [31], a reduced-representation method adapted to generating population-level genomic data in non-model species with large genome sizes. Indeed, our overarching aim was to evaluate how the temporal and spatial variation in aridity has shaped the population history and structure of *R. bechuanae*, by integrating inferences of historical changes in effective population size with GEA analyses.

Our first objective was to reconstruct past changes in effective population size in relation to palaeoclimatically inferred shifts in aridity. We expected that historical episodes of increased aridity would be associated with population declines, either due to demographic bottlenecks [32,33], or via reductions in genetic diversity caused by habitat fragmentation and decreased population connectivity. Conversely, wetter periods were expected to correspond with population expansions or increased gene flow.

112 In addition to temporal variation, spatial variation in aridity may also influence current
113 population structure and genetic diversity [12,34,35]. Therefore, our second objective
114 was to characterise the current population structure within *R. bechuanae* and assess its
115 relationship with spatial variation in aridity across the species' range. We predicted that
116 population structure would correlate with the degree of aridity, reflecting its impact on
117 gene flow and differentiation.

118 Our third objective was to assess the genomic basis of adaptation to aridity in *R.*
119 *bechuanae* using a GEA approach with the Aridity Index as a quantitative predictor. This
120 approach enabled us to identify loci associated with spatial variation in arid conditions
121 across the species' range and to identify potential biological functions that may underpin
122 survival in arid environments. We hypothesised that site frequency variation would
123 concentrate in genomic regions linked to key functions for survival in deserts such as
124 DNA repair, protein modification, osmoregulation, and carbohydrate/fat metabolism
125 [10,16].

126

127 Material & Methods

128 Samples, aridity index and RAD-seq data production

129 Individual samples of four-striped mice of the genus *Rhabdomys* were previously
130 collected in their natural environment between 2007 and 2019 by us or collaborators [25].
131 Population-level samples from *R. bechuanae* were collected specifically for the present
132 study in 2018 and 2019 (details for trapping procedures in [25]). Based on data obtained
133 for *R. pumilio* indicating minimal relatedness at this distance [36], we aimed to avoid
134 related individuals as much as possible, by selecting mice trapped at least 100 m apart,
135 except for breeding pairs (a male and a female).

136 To study fine-scale intraspecific structure, as well as perform GEA and historical effective
137 population size inferences in *R. bechuanae*, we selected a large number of individuals
138 from this species, sampled in 11 different localities at least 20 km distant from each other
139 and along the aridity gradient (*N*=230 individual samples in total, 19.82 +/- 4.72 individuals
140 per locality, **Table S1**). We estimated the aridity level of each *R.bechuanae* sampling
141 locality by computing the Aridity Index (AI) at each location, using ([38], see
142 **Supplementary Material** for details). Species identification for all sampled individuals in
143 the *Rhabdomys* species complex composed of cryptic species had previously been
144 obtained via COI sequencing or RFLP genotyping ([25] and procedure therein; **Table**
145 **S1**). DNA extractions for RAD-sequencing were performed from tail tissues preserved in
146 98% ethanol, using a 96-Well Plate Animal Genomic DNA Miniprep Kit (BioBasic Inc.).
147 Since samples from other *Rhabdomys* species and subspecies had also been collected
148 for a separate study, we proceeded with RAD-seq library preparation and sequencing for
149 *R. bechuanae* samples analysed in this study concurrently with the other *Rhabdomys*
150 samples (*N* = 411 in total). Library preparation followed the protocol of Etter et al. [39],

with modifications as described in [40] (See **Supplementary Material** for details). Sequencing was performed using a 2x150 bp paired-end Illumina NovaSeq 6000 protocol at MGX-Montpellier GenomiX Facility (Montpellier, France).

RAD-seq raw data processing and filtering

Following standard Illumina processing and quality filtering, duplicate reads resulting from PCR amplification were discarded using the program *clone_filter* implemented in the *Stacks* v. 2.66 software [41]. The resulting paired-end reads were demultiplexed and quality filtered using the *process_radtags* pipeline, also implemented within *Stacks*. Sequence reads were aligned to the reference genome using *BOWTIE* (v. 2.4.5) [42]. The reference genome used was the one of the closely related species *R. pumilio* (2.3 Gb in length; GCA_030674055.1) [43]. Before mapping, putative repeated elements were hard-masked (i.e., repeated sequences were replaced by stretch of "N") from the reference genome using the mouse repeated element Repbase Update (v. 27.06) database using *Repeatmasker* v. 4.1.3. [44]. Reads mapping to more than one reference sequence were discarded, and the maximum number of mismatches allowed was three. The resulting files were converted to .bam files and sorted using *SAMtools* v. 1.18 [45].

The reference-based *Gstacks* pipeline was applied to cluster reads into RAD loci using the Marukilow model, minimum mapping quality of 40, and alpha thresholds (for mean and variance) of 0.05 for discovering single nucleotide polymorphisms (SNPs). Using the catalogue of loci generated with *Gstacks*, the *populations* program was then used to proceed to SNP and sample filtering steps and generate SNP datasets adapted to each type of downstream analyses (see **Supplementary Material** and **Table S2** for details).

Inferences of historical changes in effective population size

To evaluate historical events that could have significantly impacted current population genetic diversity and structure, and test whether historical variations in effective population size (N_e) coincide with known past variations in the degree of aridity, we estimated historical changes in N_e using *Stairway Plot 2* v.2.1.2, which conducts multi-epoch coalescent inference to assess changes in population size over time [46,47]. We ran this analysis at two different sample scales: 1) at the *R. bechuanae* species level, pooling individual samples from all localities, and 2) at the locality level: within *R. bechuanae*, analysing two localities representing the two extremes of the Aridity Index across the sampled distribution. At each scale, we applied a downsampling procedure and run the analysis with N_{max} = 19 per locality.

The Python program *easySFS* [48] was first used to generate the folded site frequency spectrum (SFS) formatted file (v0.0.1) [49]. Since demographic inference from the SFS is particularly dependent on the presence of rare alleles (i.e., singletons and doubletons) [50], we used the SNP data set obtained without the *-mac* filter. In all cases, the samples with the least missing data were included, and then projected down, to cover the optimal

number of SNPs. For all analyses, the mutation rate was set to 5.7×10^{-9} per site per generation (based on the germline mutation rate found in *Mus musculus* in [51]), with a generation time of one year. We considered the entire allele frequency spectrum with and without singletons (See **Supplementary Material** for details).

Population structure analyses

To describe the structure of genetic diversity within *R. bechuanae*, we first computed differentiation statistics (pairwise F_{ST}) [52] for all sampled localities with sufficient sample sizes using the *vcftools* program v.0.1.16 [53]. To test isolation-by-distance and isolation-by-environment within *R. bechuanae*, simple and partial Mantel tests (R *ncf* package v. 1.3-2) [54] were conducted to test for correlations between genetic differentiation among all *R. bechuanae* localities (using $F_{ST}/(1- F_{ST})$ following [55]) and log-transformed geographic distance, as well as log-transformed environmental distance (using the Aridity Index). To characterise population structure within *R. bechuanae*, we performed a principal component analysis (PCA) and a sparse non-negative matrix factorisation (*sNMF)* analysis using the R package *LEA* v.3.14 [56]. For *sNMF* analyses, we ran 1,500 repetitions for each value of *K* ranging from 1 to 8 and chose the best *K* by evaluating cross-entropy values. Additionally, a neighbour-joining tree was generated using SNP data of these same samples with the R package *ape* (v. 5.7-1) [57].

Genotype-Environment Association analyses

To explore signals of genetic adaptation to aridity, we conducted Genotype–Environment Association (GEA) analyses. Using extensions of F_{ST}-based models, the rationale was that environmental variables distinguishing differentiated populations should be associated with allele frequency differences at loci subjected to the selective constraints they impose [22]. To test for an association between allele frequency differences between all *R. bechuanae* localities and the environmental variable of interest, the Aridity Index, which was computed for all localities using [38], we used the program *BayPass* v.2.4, which implements a method accounting for the neutral correlation of allele frequencies across populations to detect loci associated with covariables [24]. We first pooled individual genotype information from each of the 11 *R. bechuanae* localities, before running *BayPass* under the standard covariate model, which uses an Importance Sampling (IS) approximation to identify SNPs associated with the Aridity Index, and corrects for population structure by using the scaled covariance matrix Ω of population allele frequencies. Three runs (with different starting seeds) were performed to test for convergence. To determine the level of association of each SNP with the Aridity Index and identify SNPs significantly associated, we computed the empirical Bayesian p-value (eBPis) and Bayes factors (*BF*) expressed in deciban units (dB), with 10 dB being a threshold of strong association according to Jeffreys' rule [58]. Additionally, *XtX* statistics [22], which explicitly account for population structure through Ω, were outputted by the *BayPass* core model to identify SNPs with significant genetic differentiation among *R. bechuanae* populations (see **Supplementary Material** and **Figure S1** for details).

229 To construct an accurate gene annotation file, we mapped the *Mus musculus* GRCm39
230 assembly annotation [59] onto the *R. pumilio* genome using the default settings of *liftoff*
231 software version 1.6.3 [60]. All SNPs were mapped onto the annotation and we identified
232 all annotated protein-coding genes containing genotyped SNPs for functional analyses.

233 Among those, we defined outlier genes as follows: genes containing SNPs that were
234 found to be significantly associated with the Aridity Index (*BF* greater than 10 dB) were
235 defined as outlier 'association' genes, and among those, genes containing SNPs both
236 significantly associated and differentiated among populations (both with a *BF* greater
237 than 10 dB and a significant *XtX*) were defined as outlier 'association/differentiation'
238 genes. For each outlier gene, we first retrieved a full description from the Alliance of
239 Genome Resources [61], which was then summarised to obtain a meaningful set of
240 functional information. We then performed functional enrichment analyses on these two
241 lists of outlier genes using *Metascape* and the full set of annotated genes containing
242 genotyped (outlier and non-outlier) SNPs as the background. The enrichment was initially
243 considered statistically significant at $p < 0.05$. P-values were then corrected for false
244 discovery rates according to the Benjamini-Hochberg method. Only enrichments
245 exhibiting a corrected $p_{adj} < 0.05$ were considered significant.

246

247 Results

248 RAD-seq and genotype dataset

249 Across all 230 *R. bechuanae* samples analysed, an average of 2,316,057 sequence
250 reads per sample were mapped against the *R. pumilio* reference genome, after using
251 *process_radtags* to filter for quality and *clone-filter* to mask potential PCR duplicates (see
252 **Table S3** for details). The alignment rate to the repeat-masked *R. pumilio* genome
253 averaged 55.96 % across *R. bechuanae* samples. The *Gstacks* analysis resulted in an
254 average of 151,936 ± 1,779 loci per sample. RAD loci were genotyped with a mean per-
255 sample coverage of 10.94 ± 0.21 X and a mean of 227.88 ± 0.09 base pairs per locus.

256 The SNP and sample filtering procedures, used to generate the three different *R.*
257 *bechuanae* datasets for population structure analyses, inferences of past effective
258 population sizes and GEA analyses, resulted in 44,900 (population structure; GEA
259 datasets), and 147,761 (inferences of past effective population sizes dataset)) final sets
260 of SNPs (**Table S2**). After SNP and genotype quality filtering, missing data were on
261 average 9.56 ± 0.68 % per sample (**Figure S2**). A total of 154 samples were kept after
262 individual missing data filtering (< 50%) and downsampling for analysis of population
263 structure and historical changes in population effective size, while 218 samples were kept
264 after individual missing data filtering for the GEA (**Tables S2** and **S3**).

Genetic diversity and population structure in time and space

Historical changes in effective population size in *R. bechuanae*

Aridity Index (AI) estimations validated our sampling of *R. bechuanae* along an aridity gradient in South Africa: AI of sampled localities ranged from 0.0322 (Klein Pella, classified as hyper-arid) to 0.2457 (Sandveld, classified as semi-arid) (**Table S4**).

We detected no major differences in the temporal patterns of N_e between two distinct populations at the semi-arid (Sandveld) or hyper-arid (Klein Pella) extremes along the aridity gradient, nor any evidence of recent population decline (**Figure 1**, green and orange lines). The two populations showed stable and similiar N_e, from the LGM period until present, detecting only a difference between the two populations during the LGM period where Klein Pella population showed a reduction in N_e. When analysing historical changes in N_e for *R. bechuanae* as a whole, pooling data from all sampled localities, *Stairway Plot 2* inferred a distinct historical trajectory of *Ne* compared with intra-population inferences (**Figure 1**, red line). Similarly to population-specific inferences, a steep increase in N_e beginning around 140 thousand years ago (kya) was detected, followed by a stabilisation around 100 kya, with a median N_e estimate of 4,434,261 individuals reached around 60 kya. From ~30 kya - shortly before the Last Glacial Maximum (LGM) - the species-level trajectory departed from population-specific inference, with N_e declining gradually and then sharply from ~15 kya to the present. This trajectory was also recovered in the analysis excluding singletons (**Figure S3**). Comparison with palaeoclimatic reconstructions from the central summer rainfall zone of Southern Africa [62], revealed that this period of sharp N_e decline coincides with repeated and localised arid episodes occurring at 15 – 11, 7–5 and 1.5–0.5 kya (**Figure 1**).

Contemporary *R. bechuanae* population structure

The analysis of contemporary population structure on all retained *R. bechuanae* individuals revealed some degree of genetic clustering despite low level of intraspecific differentiation within this species (**Figure 2**, **Tables S5 and S6**). Pairwise F_{ST} estimates among *R. bechuanae* sampling localities were relatively low, spanning from 0.010 to 0.058 (**Table S5**). Combined, the Principal Component Analysis (PCA) (**Figure S4**), the *sNMF* analysis (**Figure 2A**) and the neighbour-joining tree (**Figure 2C**) did not indicate a clear-cut separation of several sampling localities or groups of localities from each other.

The *sNMF* analysis yielded similar cross-entropy values for K = 3 and 4 clusters (**Figure S5**), suggesting that these two inferences of population structure are almost equally likely (**Figure 2A**). At the lowest cross-entropy value (K=3; **Figure 2**, **Table S6**), a first *R. bechuanae* genetic cluster grouped most localities from a region corresponding to the Southern Kalahari (Lake Naute, Mariental, Tswalu, Molopo, Kolomela, Sandveld), which ranged from a hyper-arid to a semi-arid climate according to the estimated Aridity Index. Another cluster was more prevalent in the Gariep and Tussen die Riviere semi-arid

305 localities, and a third cluster was mostly found in the hyper-arid Klein Pella locality (**Figure**
306 **2B**, **Table S6**). Genetic differentiation between these groups of localities was also visible
307 on the neighbour-joining tree (**Figure 2C**). Most individuals from localities located
308 centrally in our sampling scheme (Benfontein, Kolomela Mine) did not have ancestry
309 coefficients > 85 % from one specific cluster and were instead composed of mixed
310 continuous ancestries (**Figure 2C, Table S6**), as also suggested by the PCA plot (**Figure**
311 **S4**).

312 Consistent with the gradual pattern of genetic ancestries observed across the sampled
313 localities (**Figure 2**), simple Mantel tests using genetic distances based on F_{ST} values
314 among *R. bechuanae* localities (**Table S5**) showed significant positive correlations
315 between genetic distance and geographic distance (p = 0.038), indicating isolation by
316 distance, as well as between genetic distance and the environmental variable of interest,
317 the Aridity Index (p = 0.037), indicating isolation by environment (**Figure 3**). Since we
318 found a very strong correlation between geographic and environmental distances (p
319 <0.001) and no significant partial Mantel tests (**Table S7**), these results suggest that
320 patterns of isolation by distance and by environment are closely intertwined, and that
321 genetic differentiation may be shaped by the aridity gradient, which closely mirrors the
322 geographic distance gradient.

323

324 Genetic basis of adaptation to aridity

325 Genotype-Environment Association analysis and genomic scan for differentiation

326 To take advantage of the full aridity gradient across *R. bechuanae*'s area of distribution,
327 we performed the GEA analysis at the locality level using *BayPass*. Given the above
328 correlations, we expected stringent conditions for the GEA analysis. Indeed, since
329 *BayPass* explicitly standardises population allele frequencies for their correlation
330 structure - here induced by the neutral population structure but also the isolation by
331 environment signal - we expected a high rate of false negatives when testing for an
332 association with the Aridity Index.

333 Overall, despite these very stringent conditions, 108 SNPs were found to be significantly
334 associated with the Aridity Index (significance threshold of 10 dB), distributed across all
335 genomic scaffolds (**Figure 4A**, **Table S8**). Among these, 68 were mapped within an
336 annotated coding region of the *R. pumilio* genome, corresponding to 63 outlier
337 'association' genes in total.

338 Across the three runs of the model, which gave similar results, a total of 1,682 variants
339 were identified as *XtX* outliers (**Figure 4B**). Of these, 65 SNPs displayed strong evidence
340 of association with aridity (significance threshold of 10 dB), as well as a significant *XtX*
341 value. Among them, 45 were mapped within an annotated coding region of the *R. pumilio*
342 genome, corresponding to 41 outlier 'association/differentiation' genes.

343

344 ### Functional analyses

345 These 41 'association/differentiation' outlier genes corresponded to twelve broad
346 functional physiological pathways (**Table S8**), reflecting diverse metabolic functions. In
347 **Figure 4**, we present one representative gene from each pathway, selected for showing
348 the strongest signal of association with the Aridity Index. Among the most promising
349 candidate genes were those involved in nervous system functions (e.g., *ADGRB1*,
350 *WHAMM,* see **Table S8**). Notably, Gene Ontology (GO) terms related to
351 neurotransmission showed the strongest degree of enrichment ($p_{adj} < 0.01$) among genes
352 associated with the Aridity Index, with twelve GO terms significantly enriched overall, as
353 compared to the background genome **(Figure 5, Table S9**).

354 Several genes were also linked to maintaining cellular integrity and stability of genetic or
355 protein information. These included genes associated with osmoregulation (e.g., *EFCC1*),
356 cellular response to stress (e.g., *SLC37A3*), and protein metabolism (e.g., *TMEM44*).
357 Both gene sets – those associated with aridity and those both associated and
358 differentiated – contained genes involved in the regulation of transcription (e.g., *ESRP2*,
359 see **Table S8**). Relevant GO terms were significantly enriched among genes associated
360 with aridity (**Figure 5**), and approached significance (p_{adj} = 0.063) among
361 associated/differentiated genes **(Table S9**). Furthermore, GO terms related to
362 cytoskeleton regulation and GTPase activity were also near the threshold for enrichment
363 significance in the associated/differentiated gene set (p_{adj} = 0.063) (**Table S9**), with genes
364 such as *WHAMM* or *EEFSEC* among top candidates (**Figure 4, Table S8**).

365 Finally, several genes with potential relevance for energy conservation were identified.
366 These included genes involved in lipid metabolism (e.g., *SDR9C7*), carbohydrate
367 metabolism (e.g., *GNAQ*), immunity (e.g., *STING1*), and sweet taste perception
368 (*TAS1R2*, see **Table S8**).

369 ## Discussion

370 In this study, we generated RAD-seq data, enabling us to investigate population-level
371 patterns of intraspecific genetic diversity and responses to aridity in the African four-
372 striped mouse, *Rhabdomys bechuanae*, a rodent species broadly distributed across an
373 aridity gradient in Southern Africa. By developing an original integrative approach,
374 combining genomic data with fine-scale contemporary and historical aridity data available
375 for the Southern African region and leveraging a natural aridity gradient to investigate
376 signatures of aridity-driven intraspecific divergence using GEA analyses, our study
377 provides significant insights into population-level responses to increased aridity in a
378 homeothermic rodent species. A strong multi-level filtering process, combined with one
379 of the most conservative GEA methods [63], along with the intertwined nature of
380 geographic structure and environmental variation in our dataset, yielded a small, but
381 robust set of genomic regions associated with aridity, that are further supported by
382 meaningful functional characteristics.

383 Our results suggest a recent decline in effective population size at the species level,
384 coinciding with Holocene aridification. Additionally, we found that genetic diversity within
385 the species is shaped by spatial structure of the environment, closely linked to variation
386 in aridity across its distribution range. We detected allele frequency shifts associated with
387 aridity levels, enabling identification of candidate genes and functions potentially involved
388 in adaptation to arid environments. Although this study, sampling a portion of the genome,
389 may miss other functions contributing to this response, we identified, as predicted,
390 candidate genomic regions associated with classical desert survival phenotypes,
391 including DNA repair/protein modification, osmoregulation and carbohydrate/fat
392 metabolism [10,16]. However, remarkably, the strongest functional enrichment signal
393 emerged for genes involved in neuronal transmission functions, bringing novel insights
394 into the role of neurological processes in coping with osmotic stress and resource
395 scarcity. Together, these findings support the hypothesis that genetic variation in *R.*
396 *bechuanae* reflects both its demographic history, shaped by past climatic fluctuations,
397 and the selective pressures imposed by spatial variation in aridity across its distribution
398 range [64].

399 Population structure and historical changes in effective population size 400 in response to aridification

401 Although genetic differentiation amongst *R. bechuanae* localities was relatively low
402 overall, our analyses revealed genetic clustering (three clusters) with the *sNMF* approach,
403 along with significant correlation between genetic and geographical distances. This
404 intraspecific structure was also supported by both PCA and neighbour-joining analyses.
405 The observed pattern of isolation by distance provided evidence of restricted gene flow
406 across the species' distribution range, which aligns with the patchiness of favourable
407 habitat available for *R. bechuanae* in the arid and semi-arid regions of Southern Africa
408 [25,65]. Isolation by distance results from constraints on dispersal, through habitat
409 patchiness for example, thereby promoting the population structure observed here
410 despite the species' relatively continuous regional distribution. Our findings support this
411 hypothesis through the detection of significant isolation by environment associated with
412 the Aridity Index variable, although this variable showed strong correlation with
413 geographic distance. The population structure analysis identified genetic clusters that
414 correspond to populations at the opposing ends of the aridity gradient (Klein Pella
415 representing the arid extreme, and Sandveld the semi-arid extreme), both of which have
416 relatively peripheral positions in the species' range. This spatial arrangement suggests
417 that the observed genetic structure could reflect adaptation to distinct environmental
418 conditions along the aridity gradient, which may have imposed divergent selective
419 pressures on *R. bechuanae* populations, supporting the hypothesis that aridity is a key
420 driver of population structure in this species.

421 Historical demographic events, such as range expansions, population bottlenecks or
422 divergence due to local adaptation or neutral processes, may have shaped patterns of
423 population structure and diversity in *R. bechuanae* and may be partly related to past

climatic events. This motivated us to infer historical changes in effective population size and correlate results with known past climatic events.

A marked decline in the hyper-arid locality Klein Pella around the Last Glacial Maximum was observed, suggesting that biome changes induced by the last deglaciation in Southern Africa [66] partly drove *R. bechuanae* population history. According to palaeoclimatic reconstructions from central Southern Africa, aridification took place in the *R. bechuanae* distribution range in the late Pleistocene [62]. During this time, the Southern Kalahari experienced a gradual increase in temperature [67] that may have contributed to N_e reduction in ancestral populations observed at the arid extreme, upholding our prediction of N_e decline during dry periods. In more recent times, single-locality effective population size (N_e) was largely stable and similar between 'arid' or 'semi-arid' localities. The similar effective population size trends inferred in relatively peripheral localities at each environmental ‘extreme’ in recent times can be attributed to low migrant rates [68].

Different trajectories between single-locality and species-level inferences were also observed, as a decline in N_e was inferred at the species level from 15 thousand years ago, but not in any of the localities on their own.

In this study, it seems unlikely that the recent reduction in N_e at the species level reflects changes in effective population size : these patterns could instead be consistent with a strong effect of structuration within *R. bechuanae*. Indeed, signals of past changes in population size in such inferences can often reflect stable population structure or variations in population connectivity [68-70]. However, this effect is likely a complex mix of intra- and inter-deme effects and difficult to fully analyse with SFS-based inference methods [68]. Given the coincidence between the period when a signal of reduced N_e is detected and repeated and localised arid past episodes [62], structuration and connectivity may have been influenced by fluctuation in aridity. Specifically, arid conditions and their impact on the vegetation may alter population dynamics of striped mice [71], as well as the distribution of nesting sites, concentrating them in scattered patches of appropriate vegetation [65], and lack of favourable habitat may create barriers that impede dispersal and gene flow between populations [72].

The patterns of genetic diversity emerging from aridification remain relatively unexplored in the literature; previous studies on vertebrate models report a link between past changes in aridity and genetic diversification [73], or changes in connectivity [74]. Overall, our results support the hypothesis that past climatic fluctuations in the Southern Kalahari and central South Africa may have been important factors shaping effective population size and population structure of *R. bechuanae,* although other neutral and adaptive factors could also be involved. Historical periods of increased aridity could not only have reduced population connectivity but may have also created opportunities for local adaptation if geographically separated populations experienced different environmental pressures along the aridity gradient. This dual effect of aridity, both fragmenting populations and creating divergent selective environments, provides a comprehensive explanation for the observed genetic structure in *R. bechuanae*.

466 ## Adaptive genomic variation along an aridity gradient

467 This study enabled the identification of several SNPs associated with the Aridity Index
468 variable, despite the very stringent conditions of the *BayPass* analysis due to significant
469 correlations between genetic, geographic and environmental (Aridity Index) distances.
470 While these conditions reduced the overall power of the analysis, they also strengthened
471 the reliability of the identified SNPs. Additional approaches, such as genome-wide
472 association studies based on well characterised physiological phenotypes [30], could
473 complement our findings by revealing other genomic regions underlying key traits for
474 adaptation to aridity. It is unlikely our inferences were biased by small sample sizes in
475 some localities, as GEA studies using Bayes' factor are not particularly sensitive to
476 sample sizes [75]. We only focused on SNPs mapping inside the coding region of genes,
477 hence proposing a very stringent view of the genes associated with the Aridity Index;
478 more genome regions, in linkage disequilibrium with the regions we identified, might also
479 contribute to the response to aridity, but we decided to focus on the most robust genomic
480 regions to interpret functionally our results.

481 Our hypothesis of differential selective pressures acting on *R. bechuanae* populations
482 along the aridity gradient was supported by the identification of outlier SNPs located
483 within coding regions of genes associated with important biological functions for
484 adaptation to an arid or desert lifestyle. We identified genes involved in energetic balance
485 (e.g., lipid metabolism, immunity) and osmotic balance (e.g., osmoregulation, neuronal
486 processes, body fluids regulation), as well as cellular responses to dehydration (e.g.,
487 protein metabolism, regulation of transcription and cytoskeleton). *Rhabdomys bechuanae*
488 face environmental constraints, which can involve high temperatures and resource
489 scarcity. Even in semi-arid conditions, these challenges cause a strain on homeostasis,
490 as revealed in *R. bechuanae* populations from semi-arid locations by blood markers of
491 osmoregulation, immunity and energy levels [30]. More arid conditions could cause
492 cellular stress, and exacerbate the effects of dehydration due to lack of water and rapid
493 water loss (e.g., respiratory water loss, evaporative cooling). Overall, consistent with our
494 predictions, our results suggest that various physiological mechanisms may enable *R.*
495 *bechuanae* to cope with harsh conditions and inhabit drier environments. We synthesised
496 previous findings from other species aligning with our results in **Table S8**, and discuss in
497 more detail the functions of interest for adaptation to arid conditions in light of this
498 literature review.

499 In numerous studies carried out on mammalian species, the same genes, or those
500 involved in similar functions as reported in our study, are highlighted as containing SNPs
501 varying with an aridity climatic index (e.g., Moisture index [76]), differentially expressed
502 in arid conditions (e.g., water restriction [20]), or showing signs of selection in specialist
503 organisms (e.g. [16]). Many of these gene lists identified from previous studies and in our
504 study intersect, suggesting a genetic basis for adaptation to aridity partly shared across
505 various mammals. Studies on the basis of adaptation to xeric environments show
506 convergent changes in tissue composition, gene expression as well as coding
507 sequences, with significant overlap mostly among related species [10,77].

508 Our findings align with key shared mechanisms of adaptation to low resource availability
509 observed across mammals, especially those involved in maintaining energetic balance,
510 such as lipid metabolism [10,13,78], and immune response [79,80]. Interestingly,
511 acquiring energy in a food-scarce environment may also involve diet shifts in response to
512 arid food sources, such as desert plants and insects. These shifts can be facilitated by
513 mutations in taste receptor genes, such as the sweet taste receptor identified in our study
514 or in the numbat *Myrmecobius fasciatus* [81], and bitter taste receptors in the *Peromyscus*
515 *eremicus* cactus mouse [15].

516 At the cellular level, dehydration-induced changes, such as cytoskeletal rearrangements
517 [82], may explain the observed association between aridity and genes regulating the
518 cytoskeleton (e.g., GO terms: filopodium assembly), suggesting possible adaptations to
519 hyperosmotic stress in the striped mouse. However, dehydration also affects cells through
520 protein denaturation, transcription/translation suppression, altered adhesion, increased
521 DNA breaks and protein oxidation, cell cycle arrest, and apoptosis [82-84]. Consistent
522 with these responses, SNPs associated with aridity were found in genes involved in
523 protein synthesis and modification, DNA repair, and cellular cycle regulation. Genes
524 containing outlier SNPs were also enriched for GO terms related to transcription
525 regulation. These findings align with previous results from selective sweeps studies in
526 arid-adapted rodents (e.g., *Peromyscus eremicus* [15]), and meta-analyses of metazoan
527 genomic and transcriptomic data showing selection on translation-related genes in harsh
528 environments [85]. Indeed, cellular responses to hyperosmotic stress commonly involve
529 regulating protein translation and degradation [84]. Overall, our results suggest strong
530 selection on genes mitigating thermal or hyperosmotic stress and restoring cellular
531 function after acute stress [83].

532 Arid environments impose strong selective pressures on osmoregulation mechanisms.
533 In *R. bechuanae*, we identified genes containing significantly associated outlier SNPs
534 involved in osmoregulation, many of which were also reported in other desert rodents
535 (e.g., *Allactaga sibirica*, *Dipus sagitta*, *Meriones meridianus* and *Phodopus roborovskii*
536 [16], **Table S8**). Water balance in mammals is regulated by brain-body communication
537 mediated by central nervous system structures that exhibit marked plasticity in response
538 to environmental cues in arid-adapted species [86,87]. This may help explain the strong
539 enrichment of GO terms related to nervous system functions, such as nervous system
540 development and neurotransmitter regulation, which emerged as the most significant
541 signal in our study. Although neurological functions are occasionally noted in aridity
542 adaptation studies (e.g., [13,75,88,89]), they are rarely emphasised as key survival
543 factors under arid conditions. When mentioned, they are typically discussed in the context
544 of selective brain cooling in large, heat-adapted mammals [90,91]. Neurotransmitters,
545 essential for maintaining homeostasis [92], regulate blood flow – a critical function in arid
546 environments where dehydration can restrict organ perfusion [93]. Our identification of
547 outlier genes involved in heart rate regulation and vasculature development - also
548 differentially expressed in other desert-adapted rodents such as *Abrothrix olivacea* [94],
549 supports this. Together, these results support the involvement of a previously
550 underappreciated functional category – neurological pathways - in the genomic response

to aridity, offering new directions for research in the evolutionary physiology of aridity adaptation.

Conclusion

This study sheds new light on how historical climatic fluctuations and environmental gradients shaped population structure and genetically based responses to aridity in *Rhabdomys bechuanae*, a rodent species inhabiting some of the most arid regions of Southern Africa. By integrating genomic data with fine-scale environmental information, we uncovered signatures of population differentiation and polygenic adaptation along an aridity gradient, involving functions linked to osmoregulation, energetic balance, and cellular stress responses. The list of candidate genes and biological functions provides a valuable foundation for further research on the molecular mechanisms of adaptation to arid environments. While we focused on coding regions to identify genes and functions associated with the Aridity Index, future work incorporating transcriptomic and non-coding genomic data will be essential to assess the role of gene expression regulation in this species' response to increasing aridity and to disentangle the relative contributions of coding versus regulatory variation. Moreover, although this study revealed genetically based signatures of adaptation, avoidance behaviours and plastic physiological responses likely play a critical role in coping with environmental stress imposed by seasonal variation in aridity [89]. Future studies manipulating water and food availability, will be key to elucidating the interplay between phenotypic plasticity and local adaptation in shaping species' responses to arid conditions.

References

1. Wiens JJ. Climate-related local extinctions are already widespread among plant and animal species. PLoS Biol. 2016 Dec;14(12):e2001104. (https://doi.org/10.1371/journal.pbio.2001104)

2. Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, et al. European phenological response to climate change matches the warming pattern. Glob Chang Biol. 2006 Oct;12(10):1969–76. (https://doi.org/10.1111/j.1365-2486.2006.01193.x)

3. Chen IC, Hill JK, Ohlemuller R, Roy DB, Thomas CD. Rapid Range Shifts of Species Associated with High Levels of Climate Warming. Science. 2011;333(6045):1024–6. (https://doi.org/10.1126/science.1206432)

4. Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO. Climate extremes: Observations, modeling, and impacts. Science. 2000 Sept;289(5487):2068–74. (https://doi.org/10.1126/science.289.5487.2068)

5. Naumann G, Alfieri L, Wyser K, Mentaschi L, Betts RA, Carrao H, et al. Global changes in drought conditions under different levels of warming. Geophysical Research Letters. 2018;45(7):3285–96. (https://doi.org/10.1002/2017GL076521)

6. Engelbrecht FA, Steinkopf J, Padavatan J, Midgley GF. Projections of future climate change in southern Africa and the potential for regional tipping points. In: Ecological Studies. Cham: Springer International Publishing; 2024. p. 169–90. (Ecological studies: analysis and synthesis. Berlin, Heidelberg, New York NY).

7. Skelly DK, Joseph LN, Possingham HP, Freidenburg LK, Farrugia TJ, Kinnison MT, et al. Evolutionary responses to climate change. Conserv Biol. 2007 Oct;21(5):1353–5.

8. Willmer P, Stone G, Johnston I. Environmental Physiology of Animals. John Wiley & Sons; 2009.

9. Rymer TL, Pillay N, Schradin C. Resilience to droughts in mammals: A conceptual framework for estimating vulnerability of a single species. Q Rev Biol. 2016 June;91(2):133–76. (https://doi.org/10.1086/686810)

10. Rocha JL, Godinho R, Brito JC, Nielsen R. Life in deserts: The genetic basis of mammalian desert adaptation. Trends Ecol Evol. 2021 July;36(7):637–50. (https://doi.org/10.1016/j.tree.2021.03.007)

11. Malaspinas AS, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, et al. A genomic history of Aboriginal Australia. Nature. 2016 Oct;538(7624):207–14. (https://doi.org/10.1038/nature18299)

12. Rocha JL, Silva P, Santos N, Nakamura M, Afonso S, Qninba A, et al. North African fox genomes show signatures of repeated introgression and adaptation to life in deserts. Nat Ecol Evol. 2023 Aug;7(8):1267–86. (https://doi.org/10.1038/s41559-023-02094-w)

13. Kim ES, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, et al. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity (Edinb). 2016 Mar;116(3):255–64. (https://doi.org/10.1038/hdy.2015.94)

14. Ababaikeri B, Abduriyim S, Tohetahong Y, Mamat T, Ahmat A, Halik M. Whole-Genome Sequencing of Tarim Red Deer (Cervus Elaphus Yarkandensis) Reveals Demographic History and Adaptations to an Arid-Desert Environment. Frontiers in Zoology. 2020 Oct;17(1):31. (https://doi.org/10.1186/s12983-020-00379-5)

15. Tigano A, Colella JP, MacManes MD. Comparative and population genomics approaches reveal the basis of adaptation to deserts in a small rodent. Mol Ecol. 2020 Apr;29(7):1300–14. (https://doi.org/10.1111/mec.15401)

16. Cheng J, Peng X, Li H, Feijó A, Xia L, Shenbrot GI, et al. Similar adaptative mechanism but divergent demographic history of four sympatric desert rodents in Eurasian inland. Commun Biol. 2023 Jan;6(1):33. (https://doi.org/10.1038/s42003-023-04415-y)

17. Bittner N, Mack KL, Nachman MW. Gene expression plasticity and desert adaptation in house mice. Evolution. 2021;75:1477–91. (https://doi.org/10.1111/evo.14172)

18. Bridges TE, James NV. THE HYPOTHALAMO-NEUROHYPOPHYSIAL SYSTEM OF NATIVE AUSTRALIAN DESERT RODENTS. THE VASOPRESSIN AND OXYTOCIN CONTENTS OF HYPOTHALAMUS AND POSTERIOR PITUITARY OF NOTOMYS ALEXIS AND PSEUDOMYS AUSTRALIS COMPARED WITH THOSE OF THE LABORATORY RAT AND MOUSE IN DIFFERENT STATES OF WATER BALANCE. Australian Journal of Experimental Biology and Medical Science. 1982;60(3):265-83. (https://doi.org/10.1038/icb.1982.31)

19. MacManes MD, Eisen MB. Characterization of the transcriptome, nucleotide sequence polymorphism, and natural selection in the desert adapted mousePeromyscus eremicus. PeerJ. 2014 Oct;2:e642. (https://doi.org/10.7717/peerj.642)

20. Marra NJ, Romero A, DeWoody JA. Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed byRNA-seq. Mol Ecol. 2014 June;23(11):2699–711. (https://doi.org/10.1111/mec.12764)

21. Coop G, Witonsky D, Di Rienzo A, Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. Genetics. 2010 Aug;185(4):1411–23. (https://doi.org/10.1534/genetics.110.114819)

22. Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 2013 July;30(7):1687–99. (https://doi.org/10.1093/molbev/mst063)

23. Gautier M. Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics. 2015 Dec;201(4):1555–79. (https://doi.org/10.1534/genetics.115.181453)

24. Berdugo M, Vidiella B, Solé RV, Maestre FT. Ecological Mechanisms Underlying Aridity Thresholds in Global Drylands. Functional Ecology. 2021;36(1):4–23. (https://doi.org/10.1111/1365-2435.13962)

25. Ganem G, Dufour CMS, Avenant NL, Caminade P, Eiseb SJ, Tougard C, et al. An update on the distribution and diversification of Rhabdomys sp. (Muridae, Rodentia). J Vertebr Biol. 2020 July;69(2):1. (https://doi.org/10.25225/jvb.20013)

26. du Toit N, Jansen van Vuuren B, Matthee S, Matthee CA. Biome specificity of distinct genetic lineages within the four-striped mouse Rhabdomys pumilio (Rodentia: Muridae) from southern Africa with implications for taxonomy. Mol Phylogenet Evol. 2012 Oct;65(1):75–86. (https://doi.org/10.1016/j.ympev.2012.05.036)

27. Meynard CN, Pillay N, Perrigault M, Caminade P, Ganem G. Evidence of environmental niche differentiation in the striped mouse (Rhabdomyssp.): inference from its current distribution in southern Africa. Ecol Evol. 2012 May;2(5):1008–23. (https://doi.org/10.1002/ece3.219)

28. Dupont LM, Rommerskirchen F, Mollenhauer G, Schefuß E. Miocene to Pliocene changes in South African hydrology and vegetation in relation to the expansion of C4 plants. Earth and Planetary Science Letters. 2013;375:408–17. (https://doi.org/10.1016/j.epsl.2013.06.005)

29. Dufour CMS, Pillay N, Avenant N, Watson J, Loire E, Ganem G. Habitat characteristics and species interference influence space use and nest-site occupancy: implications for social variation in two sister species. Oikos. 2019 Apr;128(4):503–16. (https://doi.org/10.1111/oik.05357)

30. Keilani H, Avenant N, Caminade P, Pillay N, Ganem G. Negative impact of mild arid conditions in natural rodent populations revealed using markers of physiological condition in natura. Peer Community Journal. 2025; (https://doi.org/10.24072/pcjournal.523)

31. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE. 2008;3(10). (https://doi.org/10.1371/journal.pone.0003376)

32. Benning JW, Faulkner AE, Moeller DA. Rapid evolution during climate change: demographic and genetic constraints on adaptation to severe drought. Proceedings of The Royal Society B: Biological Sciences. 1998;290. (https://doi.org/10.1098/rspb.2023.0336)

33. Ventura-Rojas PD, González-Romero A, Moreno CE, Sosa VJ. Effect of rainfall, temperature and climate change on the ecology of the rodents of arid zones: a review. Mamm Rev. 2025 Apr;55(2). (https://doi.org/10.1111/mam.12372)

34. Gomulkiewicz R, Houle D. Demographic and genetic constraints on evolution. Am Nat. 2009 Dec;174(6):E218–29. (https://doi.org/10.1086/645086)

35. Sarabia C, vonHoldt B, Larrasoaña JC, Uríos V, Leonard JA. Pleistocene climate fluctuations drove demographic history of African golden wolves (Canis lupaster). Mol Ecol. 2021 Dec;30(23):6101–20. (https://doi.org/10.1111/mec.15784)

36. Solmsen N, Johannesen J, Schradin C. Highly asymmetric fine-scale genetic structure between sexes of African striped mice and indication for condition dependent alternative male dispersal tactics. Mol Ecol. 2011 Apr;20(8):1624–34. (https://doi.org/10.1111/j.1365-294X.2011.05042.x)

38. Zomer RJ, Xu J, Trabucco A. Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. Sci Data. 2022 July;9(1):409. (https://doi.org/10.1038/s41597-022-01493-1)

39. Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: Methods in Molecular Biology.

692 Totowa, NJ: Humana Press; 2012. p. 157–78. (Methods in molecular biology (Clifton, N.J.)).
693 (https://doi.org/10.1007/978-1-61779-228-1_9)

694 40. Cruaud A, Gautier M, Galan M, Foucaud J, Sauné L, Genson G, et al. Empirical assessment
695 of RAD sequencing for interspecific phylogeny. Mol Biol Evol. 2014 May;31(5):1272–4.
696 (https://doi.org/10.1093/molbev/msu063)

697 41. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set
698 for population genomics. Molecular Ecology. 2013;22(11):3124–40.
699 (https://doi.org/10.1111/mec.12354)

700 42. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of
701 short DNA sequences to the human genome. Genome Biol. 2009 Mar;10(3):R25.
702 (https://doi.org/10.1186/gb-2009-10-3-r25)

703 43. Richardson R, Feigin CY, Bano-Otalora B, Johnson MR, Allen AE, Park J, et al. The
704 genomic basis of temporal niche evolution in a diurnal rodent. Curr Biol. 2023 Aug;33(15):3289-
705 3298.e6. (https://doi.org/10.1016/j.cub.2023.06.068)

706 44. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. 2015

707 45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. Genome Project Data
708 Processing Subgroup, The Sequence alignment/map (SAM) format and SAMtools.
709 Bioinformatics. 2009;(16):2078–9. (https://doi.org/10.1093/bioinformatics/btp352)

710 46. Liu X, Fu YX. Exploring population size changes using SNP frequency spectra. Nat Genet.
711 2015 May;47(5):555–9. (https://doi.org/10.1038/ng.3254)

712 47. Liu X, Fu YX. Stairway Plot 2: demographic history inference with folded SNP frequency
713 spectra. Genome Biol. 2020 Nov;21(1):280. (https://doi.org/10.1186/s13059-020-02196-9)

714 48. (https://github.com/isaacovercast/easySFS)

715 49. Coffman AJ, Hsieh PH, Gravel S, Gutenkunst RN. Computationally Efficient Composite
716 Likelihood Statistics for Demographic Inference. Molecular Biology and Evolution. 2016
717 Feb;33(2):591–93. (https://doi.org/10.1093/molbev/msv255)

718 50. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint
719 demographic history of multiple populations from multidimensional SNP frequency data. PLoS
720 Genet. 2009 Oct;5(10):e1000695. (https://doi.org/10.1371/journal.pgen.1000695)

721 51. Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. Differences between germline and
722 somatic mutation rates in humans and mice. Nat Commun. 2017 May;8(1).
723 (https://doi.org/10.1038/ncomms15183)

724 52. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure.
725 evolution. 1984;1358-70. (https://doi.org/10.2307/2408641)

726 53. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA. & 1000 Genomes
727 Project Analysis Group. Bioinformatics. 2011;27(15):2156–8.

728 54. Bjornstad ON, Bjornstad MON. Package `ncf. Spatial nonparametric covariance functions.
729 2016;5:1–1.

55. Rousset F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics. 1997 Apr;145(4):1219–28. (https://doi.org/10.1093/genetics/145.4.1219)

56. Frichot E, François O. LEA: An R package for landscape and ecological association studies. Methods in ecology and evolution. 2015;6(8):925–9. (https://doi.org/10.1111/2041-210X.12382)

57. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019 Feb;35(3):526–28. (https://doi.org/10.1093/bioinformatics/bty633)

58. Jeffreys H. Oxford Classic Texts in the Physical Sciences. In: Theory of Probability. Oxford: Oxford Univ. Press; 1961.

59. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002 Dec;420(6915):520–62.

60. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2021 Jun;37(12):1639–1643. (https://doi.org/10.1093/bioinformatics/btaa1016)

61. Alliance of Genome Resources Consortium. Updates to the Alliance of Genome Resources central infrastructure. Genetics. 2024 May;227(1). (https://doi.org/10.1093/genetics/iyae049)

62. Chevalier M, Chase BM. Determining the drivers of long-term aridity variability: a southern African case study. Journal of Quaternary Science. 2016;31(2):143–51. (https://doi.org/10.1002/jqs.2850)

63. Ahrens CW, Jordan R, Bragg J, Harrison PA, Hopley T, Bothwell H, Murray K, et al. Regarding the F-Word: The Effects of Data Filtering on Inferred Genotype-Environment Associations. Molecular Ecology Resources. 2021;21(5):1460–74. (https://doi.org/10.1111/1755-0998.13351)

64. Kolbe JJ, Leal M, Schoener TW, Spiller DA, Losos JB. Founder effects persist despite adaptive differentiation: A field experiment with lizards. Science. 2012 Mar;335(6072):1086–9. (https://doi.org/10.1126/science.1209566)

65. Dufour CMS, Meynard C, Watson J, Rioux C, Benhamou S, Perez J, et al. Space use variation in co-occurring sister species: Response to environmental variation or competition? PLoS One. 2015 Feb;10(2):e0117750. (https://doi.org/10.1371/journal.pone.0117750)

66. Zhao X, Lu X, Meadows M, Dupont L, Mao L, Xu Y, et al. Biome variability in southernmost Africa since the last deglaciation recorded in marine sediments. Quaternary Science Reviews. 2024;343. (https://doi.org/10.1016/j.quascirev.2024.108912)

67. Lukich V, Ecker M. Pleistocene environments in the southern Kalahari of South Africa. Quat Int. 2022 Mar;614:50–8. (https://doi.org/10.1016/j.quaint.2021.03.008)

68. Lesturgie P, Planes S, Mona S. Coalescence times, life history traits and conservation concerns: An example from four coastal shark species from the Indo-Pacific. Molecular Ecology Resources 2022;22(2) :554-66. (https://doi.org/10.1111/1755-0998.13487)

69. Heller R, Chikhi L, Siegismund HR. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. PloS one. 2013 8(5):e62992. (https://doi.org/10.1371/journal.pone.0062992)

70. Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L. On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference?. Heredity. 2016 116(4) :362-71.

71. Nater CR, van Benthem KJ, Canale CI, Schradin C, Ozgul A. Density feedbacks mediate effects of environmental change on population dynamics of a semidesert rodent. J Anim Ecol. 2018 Nov;87(6):1534–46. (https://doi.org/10.1111/1365-2656.12888)

72. Ota H. Geographic patterns of endemism and speciation in amphibians and reptiles of the Ryukyu Archipelago, Japan, with special reference to their paleogeographical implications. Researches on Population Ecology. 1998;40:189–204. (https://doi.org/10.1007/BF02763404)

73. Garcia JT, Alda F, Terraube J, Mougeot F, Sternalski A, Bretagnolle V, et al. Demographic history, genetic structure and gene flow in a steppe-associated raptor species. BMC Evol Biol. 2011;11:333. (https://doi.org/10.1186/1471-2148-11-333)

74. Merchant HN, Ivanova A, Hart DW, García C, Bennett NC, Portugal SJ, et al. Patterns of genetic diversity and gene flow associated with an aridity gradient in populations of common mole-rats, Cryptomys hottentotus hottentotus. Genome Biology and Evolution. 2024;16(7). (https://doi.org/10.1093/gbe/evae144)

75. Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Mol Ecol. 2018 May;27(9):2215–33. (https://doi.org/10.1111/mec.14584)

76. Flori L, Moazami-Goudarzi K, Alary V, Araba A, Boujenane I, Boushaba N, et al. A genomic map of climate adaptation in Mediterranean cattle breeds. Molecular ecology. 2019;28(5):1009–29. (https://doi.org/10.1111/mec.15004)

77. Chalopin D, Rey C, Ganofsky J, Blin J, Chevret P. Convergent transcriptomic and genomic adaptation in xeric rodents. Mouginot M, Boussau B, Pantalacci S, Sémon M, editors. 2025. (https://doi.org/10.1101/2024.10.02.616319)

78. Sugden LA, Atkinson EG, Fischer AP, Rong S, Henn BM, Ramachandran S. Localization of adaptive variants in human genomes using averaged one-dependence estimation. Nat Commun. 2018 Feb;9(1). (https://doi.org/10.1038/s41467-018-03100-7)

79. McDade TW, Georgiev AV, Kuzawa CW. Trade-offs between acquired and innate immune defenses in humans. Evol Med Public Health. 2016;2016(1):1–16. (https://doi.org/10.1093/emph/eov033)

80. Brusch GA, Christian KA, Brown GP, Shine R, Denardo DF. Dehydration enhances innate immunity in a semiaquatic snake from the wet-dry tropics. Journal of Experimental Zoology Part A: Ecological and Integrative Physiology. 2019;331(4):245–52. (https://doi.org/10.1002/jez.2260)

81. Peel E, Silver L, Brandies P, Hayakawa T, Belov K, Hogg CJ. Genome assembly of the numbat (Myrmecobius fasciatus), the only termitivorous marsupial. Gigabyte. 2022. (https://doi.org/10.46471/gigabyte.47)

82. Burg MB, Ferraris JD, Dmitrieva NI. Cellular Response to Hyperosmotic Stresses. Physiological Reviews. 2007;87(4):1441–74. (https://doi.org/10.1152/physrev.00056.2006)

83. Kampinga HH. Thermotolerance in mammalian cells protein denaturation and aggregation, and stress proteins. J Cell Sci. 1993 Jan;104(1):11–7. (https://doi.org/10.1242/jcs.104.1.11)

84. Lamitina T, Huang CG, Strange K. Genome-wide RNAi screening identifies protein damage as a regulator of osmoprotective gene expression. Proc Natl Acad Sci U S A. 2006 Aug;103(32):12173–8. (https://doi.org/10.1073/pnas.0602987103)

85. Porcelli D, Butlin RK, Gaston KJ, Joly D, Snook RR. The environmental genomics of metazoan thermal adaptation. Heredity (Edinb). 2015 May;114(5):502–14. (https://doi.org/10.1038/hdy.2014.119)

86. Rabhi M, Ugrumov M, Polenova O, Bengelloun W, Calas A. Modifications morphofonctionnelles du système hypothalamo-hypophysaire vasopressinergique après déshydratation chronique chez la Mérione (Meriones shawi). Ann Endocrinol. 1993;54.

87. Djazouli Alim FZ, Romanova EV, Tay YL, Rahman AYBA, Chan KG, Hong KW, et al. Seasonal adaptations of the hypothalamo-neurohypophyseal system of the dromedary camel. PLoS One. 2019 June;14(6):e0216679. (https://doi.org/10.1371/journal.pone.0216679)

88. Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, et al. Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. Molecular Biology and Evolution. 2016 Oct;33(10):2576–92. (https://doi.org/10.1093/molbev/msw129)

89. Blumstein D, Macmanes M. The multi-tissue gene expression and physiological responses of water deprived Peromyscus eremicus. BMC genomics. 2024;25(1). (https://doi.org/10.1186/s12864-024-10629-z)

90. Hetem RS, Strauss WM, Fick LG, Maloney SK, Meyer LC, Fuller A, et al. Selective brain cooling in Arabian oryx (Oryx leucoryx): a physiological mechanism for coping with aridity? Journal of Experimental Biology. 2012;215(22):3917–24. (https://doi.org/10.1242/jeb.074666)

91. Strauss WM, Hetem RS, Mitchell D, Maloney SK, OaposBrien HD, Meyer LC, et al. Body water conservation through selective brain cooling by the carotid rete: a physiological feature for surviving climate change? Conservation Physiology. 2017;5. (https://doi.org/10.1093/conphys/cow078)

92. Mittal R, Debs LH, Patel AP, Nguyen D, Patel K, OaposConnor G, et al. Neurotransmitters: The critical modulators regulating gut-brain axis. Journal of cellular physiology. 2017;232(9):2359–72. (https://doi.org/10.1002/jcp.25518)

93. Trangmar SJ, González-Alonso J. New insights into the impact of dehydration on blood flow and metabolism during exercise. Exercise and sport sciences reviews. 2017;45:146–53. (https://doi.org/10.1249/JES.0000000000000109)

94. Giorello FM, Feijoo M, DaposElía G, Naya DE, Valdez L, Opazo JC, et al. An association between differential expression and genetic divergence in the Patagonian olive mouse (Abrothrix olivacea). Molecular Ecology. 2018;27(16):3274–86. (https://doi.org/10.1111/mec.14778)

848 # Figure legends

849

850 **Figure 1: Historical changes in effective population size in *R. bechuanae*.**

851 *Stairway Plot 2* analyses (with singletons) showing historical changes in effective
852 population size (N_e) over the past 130,000 years. The red line indicates the estimated
853 median N_e, for the whole species (N=129, 146 896 SNPs), the Klein Pella locality (hyper-
854 arid climate, N=14, 146 534 SNPs) is represented in orange and the Sandveld locality
855 (semi-arid climate, N=16, 145 767 SNPs) is represented in yellow-green. The lines
856 represent the median, and the buffers represent the 95% confidence interval. The grey
857 rectangle marks the period of the Last Glacial Maximum (LGM), while the yellow
858 rectangles denote recent arid episodes in the Summer rainfall region of Southern Africa
859 [67].

860

861 **Figure 2: Population structure among *R. bechuanae* samples.**

862 **A**. *sNMF* results from K=3 to K=4 for *R. bechuanae* samples in each locality.
863 Abbreviations of sampling localities are : Klein Pella (KP); Kolomela Mine (Ko); Lake
864 Naute (LN); Mariental (Ma); Molopo (Mo); Benfontein (Be); Tswalu (Ts); Sandveld (Sa);
865 Soetdoring (So); TdR (Tussen die Riviere); Ga (Gariep Dam). **B.** Map showing the
866 distributions and genetic compositions of *R. bechuanae* sampling localities. Pie diagrams
867 indicate the frequencies of genetic clusters in each locality. Base map: World Topographic
868 Map Esri Standard, Aridity Index layer was computed from a 0.5° global grid, using data
869 from Version 3 of the Global Aridity Index and Potential Evapotranspiration Database
870 [38]. **C.** Neighbour-joining tree built using *ape.* Individuals with an ancestry coefficient >
871 0.85 for one cluster are coloured according to the colour code of the corresponding
872 genetic cluster used in panel A. *R. bechuanae* individuals without an ancestry
873 coefficients > 0.85 for any genetic cluster are coloured in grey. All analyses were
874 performed using *N*=154 individuals and 44,900 SNPs.

875

876 **Figure 3**: **Correlations between genetic distance and geographic/Aridity Index**
877 **distances**.

878 Correlation plots with slopes and correlation coefficient R^2 estimates between genetic
879 distance ($F_{ST}/(1-F_{ST})$) and log(geographic distance), and between environmental distance
880 and geographic distance matrices (orange). Environmental distance is represented by
881 Euclidean distance in terms of the Aridity Index.

882

883 **Figure 4**: **Genotype-Environment Associations and genetic differentiation in *R.***
884 ***bechuanae*.**

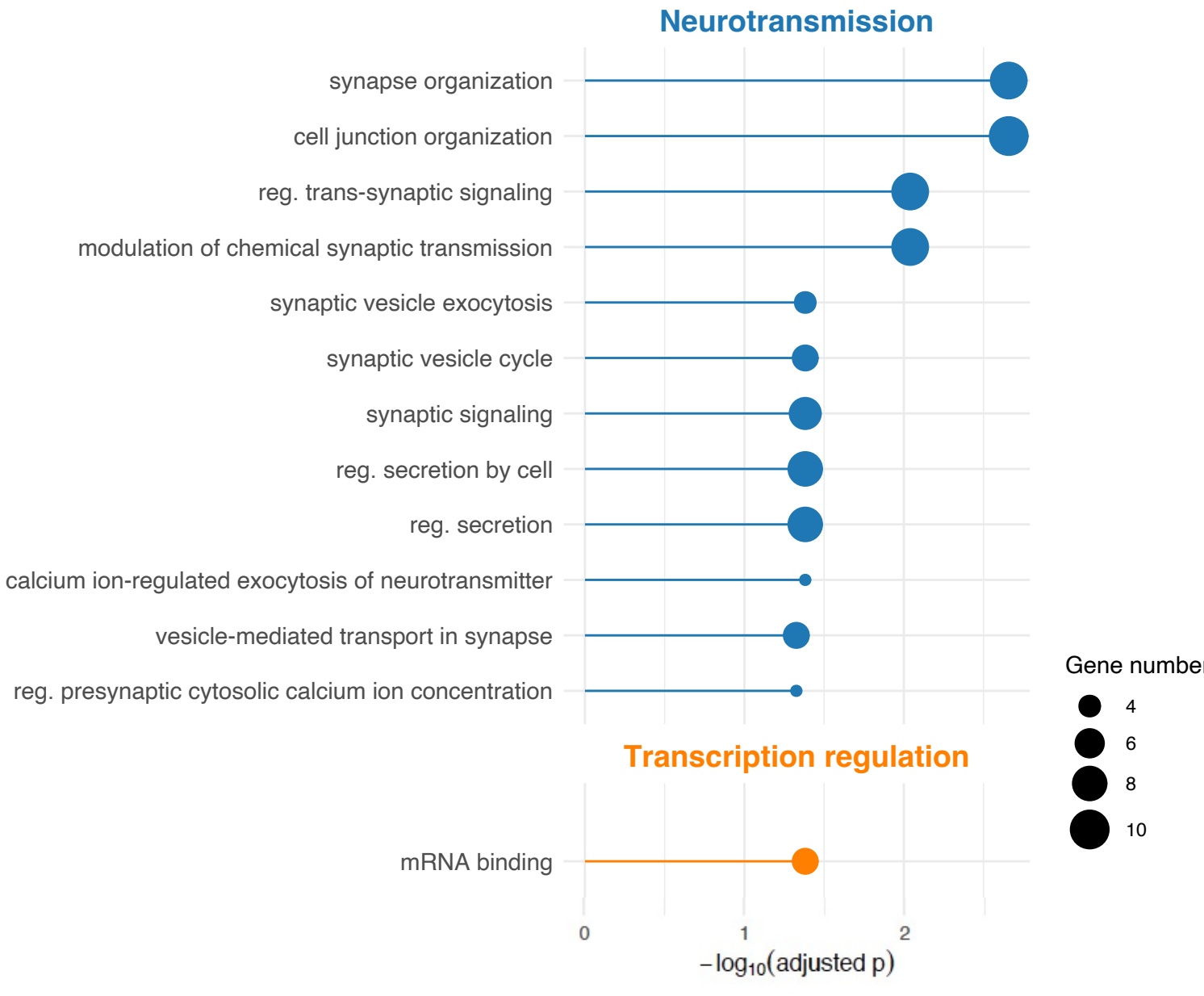
885 **A.** Manhattan plot of Bayes Factor (*BF*, in decibans) values from *BayPass* (standard
886 covariate model), showing associations between 44,900 SNPs and the Aridity Index
887 across 24 chromosome-level scaffolds (x-axis : genomic position). The dashed line marks
888 the *BF*=10 ("strong" level of association) significance threshold. Outlier SNPs are in
889 colour; others alternate black/white fill colour by scaffold). **B.** Manhattan plot of genetic
890 differentiation (*XtX* values) from *BayPass* (core model). The dashed line indicates the 5%
891 POD significance threshold (*XtX* = 14.4). Colour scheme as in A. **C**. Joint plot of *BF* and
892 *XtX* values. Thresholds are shown as dashed lines; SNPs significant for both statistics
893 are in orange. In all plots, we show the names of some genes containing significantly
894 associated and differentiated SNPs, involved in functions relevant to the response to arid
895 conditions (for each function, the gene containing the SNP with the highest BF was
896 highlighted; SNPs within these genes in red). All analyses were performed using *N*=218
897 individuals and 44,900 SNPs.

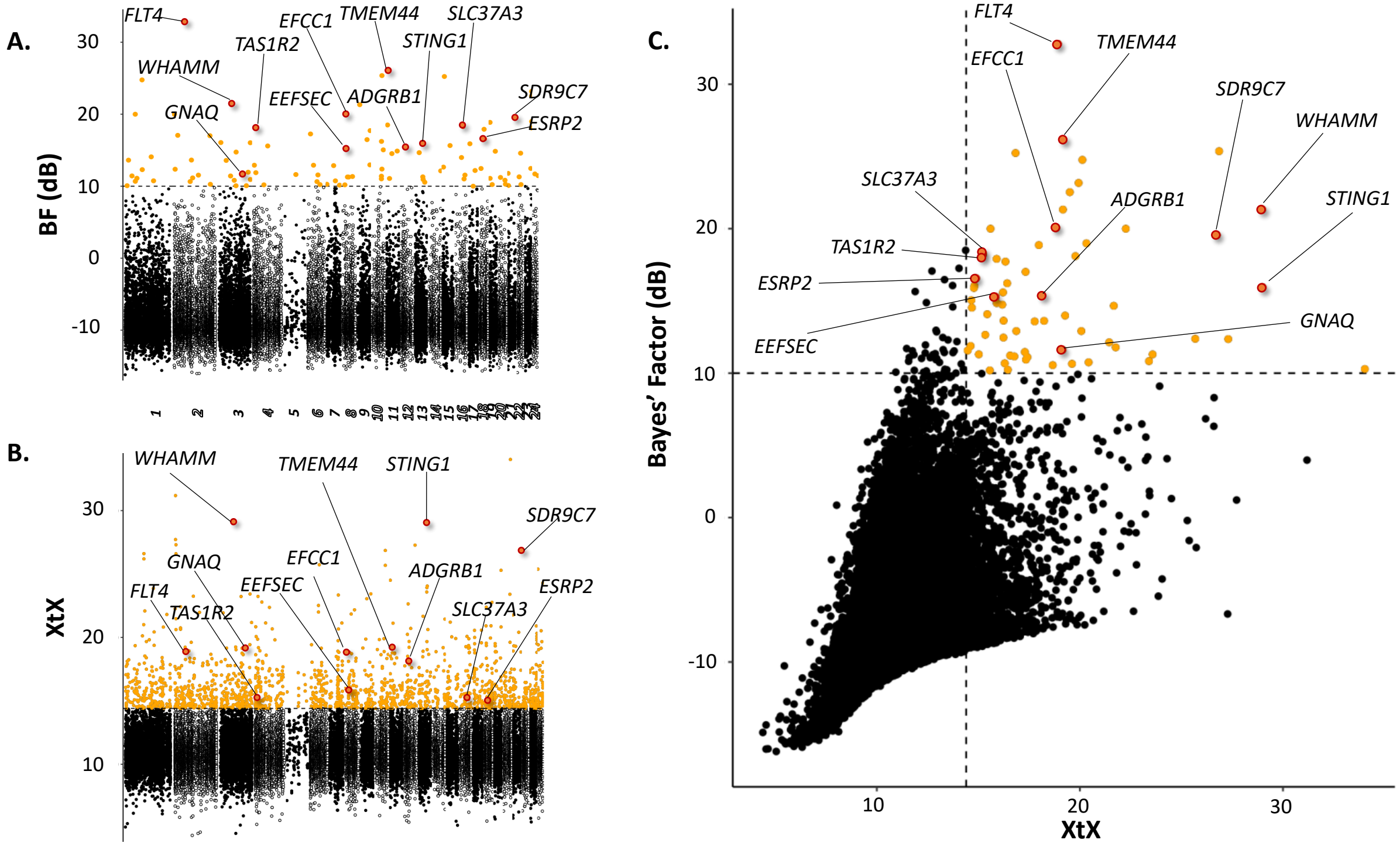
898

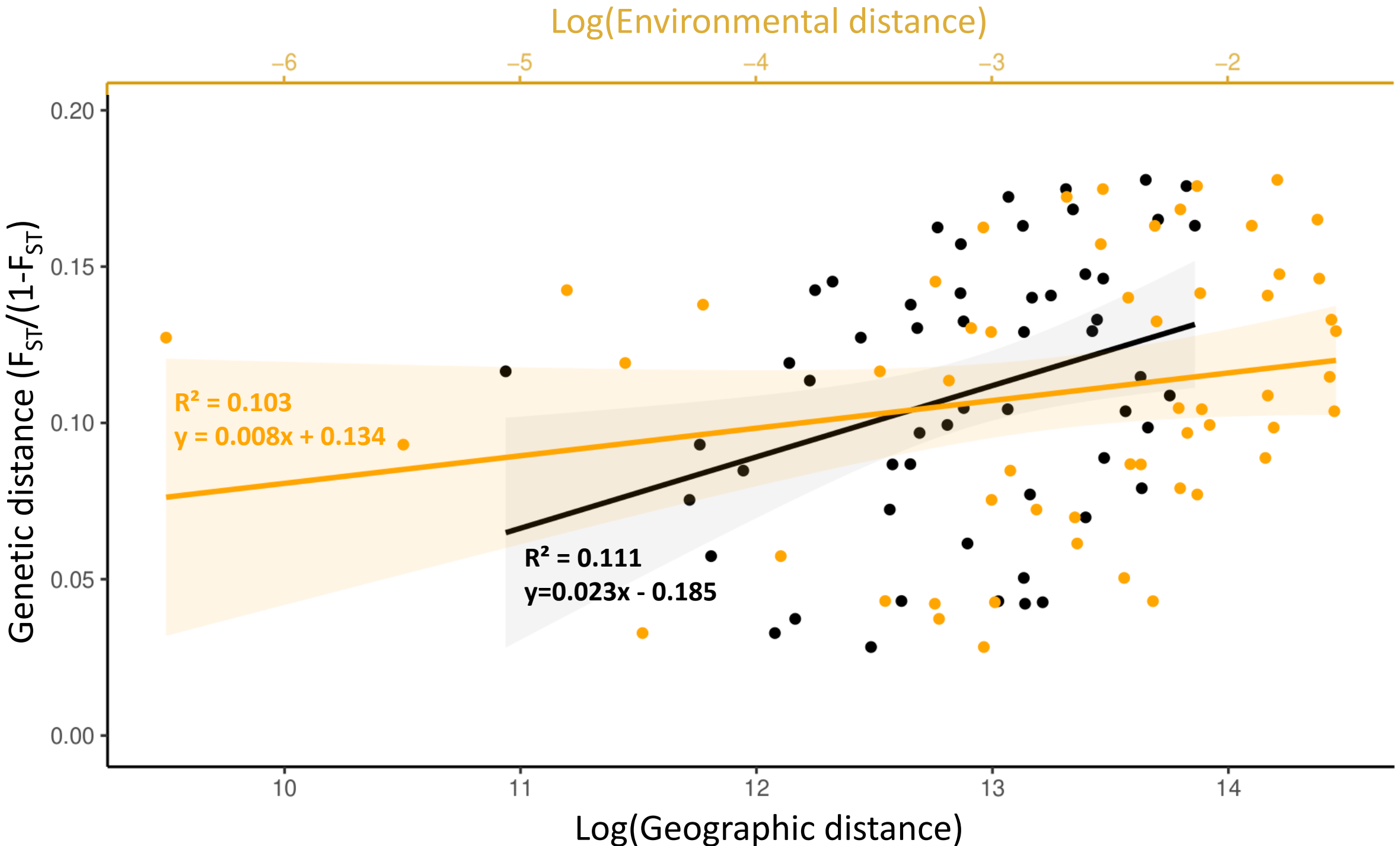
899 **Figure 5: Functional enrichment among genes significantly associated with the**
900 **Aridity Index.**

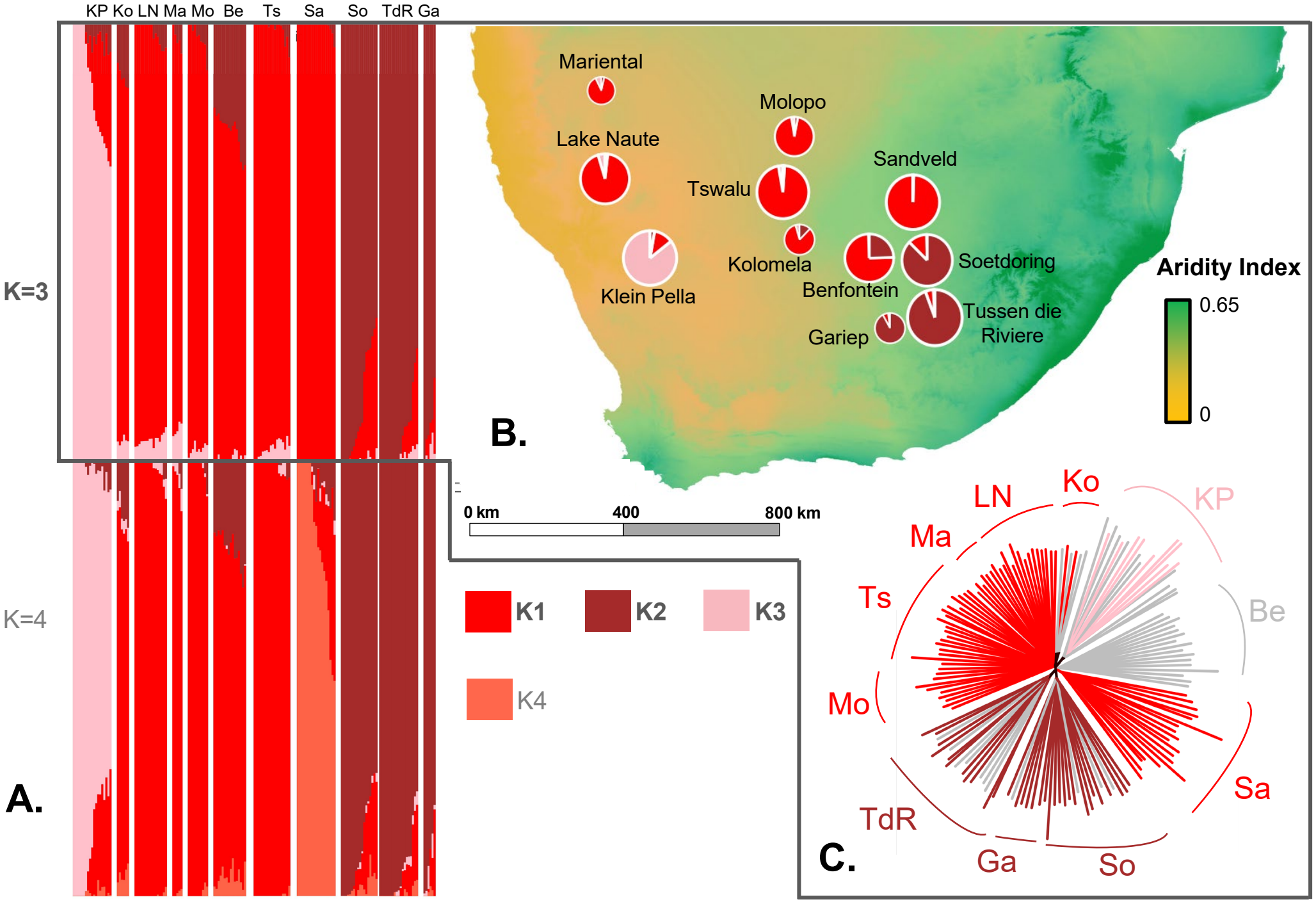
901 Abbreviated description of significantly enriched GO terms ($p_{adj} < 0.05$) are presented on
902 the left, classified into more general physiological pathways, with level of significance
903 after Benjamini-Hochberg correction reported on the x-axis. Circle size represents the
904 number of outlier genes found to be associated with each GO term.

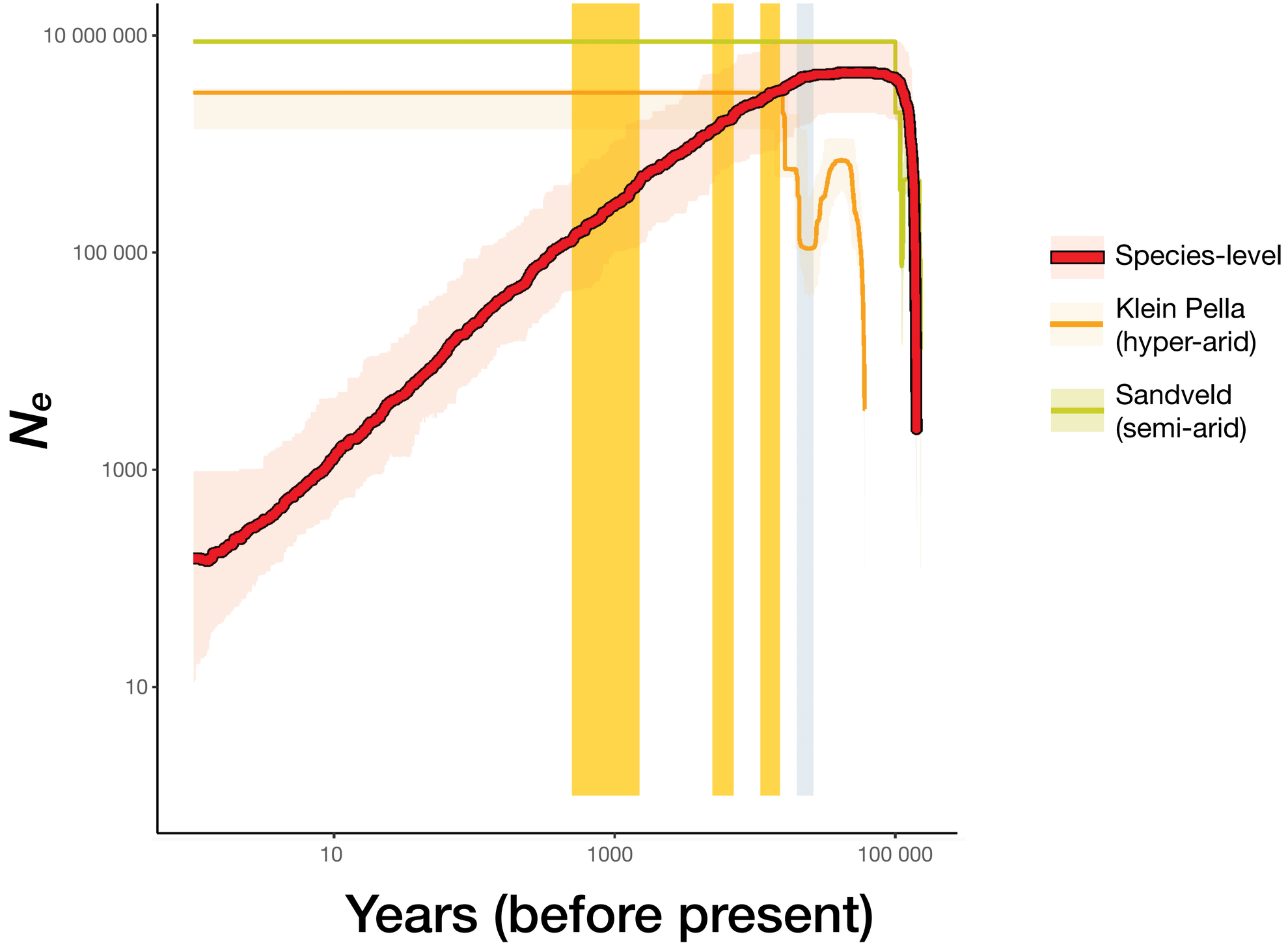
905


Neurotransmission
synapse organization
cell junction organization
reg. trans-synaptic signaling
modulation of chemical synaptic transmission
synaptic vesicle exocytosis
synaptic vesicle cycle
synaptic signaling
reg. secretion by cell
reg. secretion
calcium ion-regulated exocytosis of neurotransmitter
vesicle-mediated transport in synapse
reg. presynaptic cytosolic calcium ion concentration
Transcription regulation
mRNA binding
0
1
2
−log10(adjusted p)
Gene number
4
6
8
10


A.
BF (dB)
FLT4
EFCC1
TMEM44
SLC37A3
TAS1R2
STING1
WHAMM
EEFSEC
ADGRB1
SDR9C7
GNAQ
ESRP2
B.
XtX
WHAMM
TMEM44
STING1
SDR9C7
GNAQ
EFCC1
ADGRB1
FLT4
EEFSEC
ESRP2
TAS1R2
SLC37A3
C.
Bayes' Factor (dB)
XtX
FLT4
TMEM44
EFCC1
SDR9C7
WHAMM
SLC37A3
ADGRB1
STING1
TAS1R2
ESRP2
GNAQ
EEFSEC


Log(Environmental distance)
−6
−5
−4
−3
−2
0.20
0.15
0.10
0.05
0.00
Genetic distance ($F_{ST}/(1-F_{ST})$)
R^2 = 0.103
y = 0.008x + 0.134
R^2 = 0.111
y=0.023x - 0.185
10
11
12
13
14
Log(Geographic distance)


KP Ko LN Ma Mo Be Ts Sa So TdR Ga
K=3
K=4
A.
Mariental
Molopo
Lake Naute
Sandveld
Tswalu
Kolomela
Soetdoring
Benfontein
Klein Pella
Tussen die Riviere
Gariep
B.
Aridity Index
0.65
0
0 km
400
800 km
K1
K2
K3
K4
C.
LN
Ko
KP
Ma
Ts
Be
Mo
Sa
TdR
Ga
So


10 000 000
100 000
1000
10
Ne
10
1000
100 000
Years (before present)
Species-level
Klein Pella
(hyper-arid)
Sandveld
(semi-arid)