



Degrees of convergent evolution in rodent adaptations to arid environments

Domitille Chalopin, Carine Rey, Jeremy Ganofsky, et al.

Genome Res. 2026 36: 472-486 originally published online January 21, 2026

Access the most recent version at doi:[10.1101/gr.280089.124](https://doi.org/10.1101/gr.280089.124)

References This article cites 63 articles, 4 of which can be accessed free at:
<http://genome.cshlp.org/content/36/3/472.full.html#ref-list-1>

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Research

Degrees of convergent evolution in rodent adaptations to arid environments

Domitille Chalopin,^{1,2,5} Carine Rey,^{1,3,5} Jeremy Ganofsky,¹ Juliana Blin,¹ Pascale Chevret,⁴ Marion Mouginot,¹ Laurent Guéguen,⁴ Bastien Boussau,⁴ Sophie Pantalacci,¹ and Marie Sémon¹

¹Laboratoire de Biologie et Modélisation de la Cellule, CNRS UMR 5239, Université Claude Bernard Lyon 1, Ecole Normale Supérieure de Lyon, Université de Lyon, Lyon 69007, France; ²ImmunoConcEpT, CNRS UMR 5164, Université de Bordeaux, Bordeaux 33000, France; ³Centre International de Recherche en Infectiologie (CIRI), Inserm U1111, CNRS UMR 5308, Université Claude Bernard Lyon 1, Ecole Normale Supérieure de Lyon, Université de Lyon, Lyon 69007, France; ⁴Laboratoire de Biométrie et Biologie Evolutive, CNRS UMR 5558, Université Claude Bernard Lyon 1, Villeurbanne 69622, France

Species adapting to a similar lifestyle may undergo convergent changes in organ structure and cellular function, themselves relying or not on these convergent genetic changes. The extent of genomic convergence is thus debated and may further depend on the interplay between temporal factors, such as species relatedness or the age of the transition. Rodents have repeatedly adapted to life in arid conditions, notably with altered renal morphology and physiology. By analyzing kidney transcriptomes from 33 species, we find convergence at all examined biological levels, from the whole kidney transcriptome down to the coding sequences and expression level of individual genes. Transcriptome-level signatures reflect convergent changes in cell proportions, suggesting convergent structural adaptations of the kidney. A large proportion of genes shows convergent substitutions, but those happened in small subsets of species, showing that there are multiple genetic paths repeatedly taken in a mosaic manner. A similar mosaic signal of convergence is found comparing gene expression in species spanning the *Rodentia* order, but convergence is more widely shared at the lower level of the *Murinae* family. Therefore, we test more directly the influence of temporal factors. We observe more convergent changes when we select species independently adapted from more closely than more distantly related ancestors and when we select older transitions rather than recent transitions. Our study shows that there are many different, yet repeatedly selected, ways to adapt to aridity and that the degree of convergent evolution increases with both the age of the transitions and species relatedness.

[Supplemental material is available for this article.]

Repeated evolution, embracing parallel and convergent evolution, occurs when different lineages evolve similar traits independently. If these similar traits are underlain by the same genetic changes, the genetic basis of adaptation might be predictable. Recent genomic studies, combined with smaller-scale genetic studies, have significantly advanced our understanding of this question (e.g., Brown et al. 2019; Sackton et al. 2019; Chaturvedi et al. 2022; Zancolli et al. 2022; Corral-Lopez et al. 2024; Pollard et al. 2024). They highlighted a large variability in the degree of genomic convergence, which may be influenced by several factors. In particular, depending on the study design, time may have contradictory influences on the probability of gene reuse and the expected amount of genomic convergence (Bohutínská and Peichel 2024). As time since speciation increases, species accumulate genomic and functional divergence, so that when adaptation to a common lifestyle occurs, the probability of gene reuse should be reduced, which should result in lower amounts of genomic convergence. However, as time since transition to a given lifestyle increases, adaptation may encompass more and more phenotypic traits, increasing chances that some are shared between different lineages

and thereby increasing the probability of gene reuse and the amount of genomic convergence. To our knowledge, we so far lacked study designs that could quantify genomic convergence accounting for both species relatedness and age of transition.

Rodentia is the most diversified order of mammals with living representatives spanning 70 million years (MY) of evolution. The subfamily Murinae originated about 11.2 MY ago and today represents ~20% of all rodents (Aghová et al. 2018). In these two nested radiations, repeated transitions to arid environments have occurred, with examples of ancient adaptations as well as more recent ones. This provides a unique opportunity to quantify genomic convergence while controlling for species relatedness and age of transition.

Life in arid environments has been studied in a variety of mammalian clades. Multiple adaptations enable species to cope with temperature and seasonal unpredictability and with challenges to food and water availability and quality (Schwimmer and Haim 2009). They have motivated a rapidly growing area of research in genomics (Rocha et al. 2021). Studies include the comparison of renal gene expression in a few species (MacManes and Eisen 2014; Marra et al. 2014; Giorello et al. 2018; Bittner et al. 2022), dehydration experiments to study the plasticity of gene expression (Kim and Shin 2016; Bittner et al. 2021; Blumstein and

⁵These authors contributed equally to this work.

Corresponding authors: domitille.chalopin-fillot@u-bordeaux.fr, sophie.pantalacci@ens-lyon.fr, marie.semon@ens-lyon.fr

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.280089.124>. Freely available online through the *Genome Research* Open Access option.

© 2026 Chalopin et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

MacManes 2023), genomic analyses (Cheng et al. 2023; Peng et al. 2023), and population genomic analyses (Tigano et al. 2020; Rocha et al. 2023).

Most of the studies of transitions to arid environments have been performed in rodents. Many rodent species adapted to arid environments have acquired behavioral and physiological adaptations linked to metabolism and water retention (Rocha et al. 2021), including modified kidneys capable of producing very concentrated urine (Bankir and de Rouffignac 1985). A recent study analyzed gene expression changes and genes under positive selection in three independent transitions to desert life in rodents (Bittner et al. 2022). They discovered many idiosyncratic changes but also shared changes in genes of interest known to be involved in osmoregulation and kidney function. Overall, genes involved in lipid metabolism, responses to insulin signaling and diabetes, stress responses, the endocrine system, the arachidonic acid metabolic pathway, and water transport have all been found to be involved in the adaptation of rodents to arid environments (Giorello et al. 2018; Bittner et al. 2022; Yuan et al. 2025). This suggests that individual coding sequences and gene expression levels have undergone repeated adaptation in relation to living in arid environments, and offers a solid foundation to better characterize the factors that influence the amount of genomic and transcriptomic convergence.

Here we studied the effects of timescale on repeated genomic and transcriptomic evolution associated with transition to arid environments (Fig. 1A). First, we quantified signals of repeated evolution at two timescales, corresponding to the order Rodentia and the family Murinae. We traced signals in kidney transcriptomes at different levels: whole-transcriptome variation, sets of covarying genes, individual gene expression levels. We also tested the influence of cell composition, which has a profound impact on transcriptomic variation (Pantalacci et al. 2017), and we looked for repeated amino acid evolution in coding sequences of kidney-expressed genes. Second, we evaluated the influence of species relatedness and of the age of the transitions on these signals (Fig. 1B). For this, we applied a variety of methods (Fig. 1C).

Results

A multipurpose data set to study convergent evolution

Our goal was to study convergence at two different phylogenetic scales, Rodents and Murinae, bringing together kidney expression and coding sequences data for the kidney. We combined publicly available data (cDNA from published genomes and kidney RNA-seq data) with newly generated RNA-seq data to obtain coding sequence and expression data in a variety of rodents, ensuring a reasonable overlap between the two types of data (12 families, 48 species out of which 33 species with expression data) (phylogeny built from our assemblies in Fig. 2; Supplemental Data S1). Most of our sampling effort was meant to increase the number of *Murinae* species taken from the wild, but we also included species that are more easily obtained from laboratory colonies or commercial strains (Supplemental Tables S1–S3; Supplemental Methods).

We associated each species with a biological status corresponding to its arid (orange) or mesic (green) environment. We determined its geographical distribution area and extracted the corresponding bioclimatic variables (Methods) (Supplemental Table S4; Supplemental Methods). Because an annual average pluviometry can hide large differences between seasons, we decided to use the precipitation of the driest quarter of the year and an envi-

ronment with <40 mm was defined as arid. We annotated 28 species living in arid environments and presumably adapted to arid environments (Fig. 2). For convenience, below we refer to them as “arid species,” whereas other species are named “mesic species.”

To assess the independence of the transitions to arid environments and date them, we used a semiautomated method, enlarging our annotation to 1898 species. Then we estimated ancestral states along a published rodent phylogenetic tree containing 2260 rodent species (Fabre et al. 2012; Supplemental Methods; Supplemental Data S2–S4; Supplemental Tables S1, S4; Supplemental Fig. S1), which was congruent with our phylogenomic tree presented in Figure 2. For species in our data set and their relatives, we further manually curated the transitions based on additional information from specific literature on the concerned species (i.e., more complete phylogenies by specialists of the considered group considering phylogeography) and bracketed them to distinguish recent from older transitions (minimum dates in Fig. 2, intervals and bibliography in Supplemental Fig. S3; Supplemental Table S5; Supplemental Methods).

This provided two data sets for expression analyses, so-called Total_expr and Murinae_expr, which encompassed 33 and 11 species, respectively (Figs. 2, 3A,E). Not all species for which RNA-seq data were available were selected for these data sets, as the selection needed to balance the many requirements for gene expression analysis (reproducible kidney dissection, RNA quality and batch effect, number of replicates; available expression data for an out-group mesic species or a closely related species; minimizing overrepresentation of a genus; keeping a reasonable balance between mesic and arid species). In these final data sets, nine selected species had no replicates. They were nevertheless selected because (1) a closely related species from the same genus could serve as biological pseudoreplicates for the considered environmental transition, and/or (2) arid and corresponding mesic species were all dissected and sequenced by us at the same time, ensuring minimal confounding effects.

In contrast, for coding sequence analysis, we maximized phylogenetic power by working on the largest possible data sets, named Total_seq and Murinae_seq, which encompassed respectively all the rodent species (48) and all the Murinae species (14).

Higher degree of gene expression convergence at gene, module, and whole-transcriptome level in the *Murinae* subdata set compared with the total data set

We investigated whether differences in the whole transcriptome that were recorded agnostically by principal component analysis (PCA) or by coexpression module analysis carried a signal associated with the mesic/arid environment. Next, we investigated the possibility of identifying genes that exhibit differential expression in relation to the environment.

First, we analyzed broad patterns of expression differences in a PCA. In the Total_expr data set, the difference between arid and mesic individuals only becomes apparent when the phylogeny is accounted for (Methods) (Fig. 3B; Supplemental Fig. S4). We found that the fifth principal component (PC), which accounted for only 4% of the variation, only partially separates arid from mesic species. We verified that this was not an artifact of technical batch effects (Supplemental Fig. S5; Supplemental Table S3). In the Murinae_expr data set, the first component carries most of the separation between arid and mesic individuals and accounts for a large proportion of the variation (25%) (Fig. 3F). However, on the first two PCs, three individuals of the arid species *Mus*

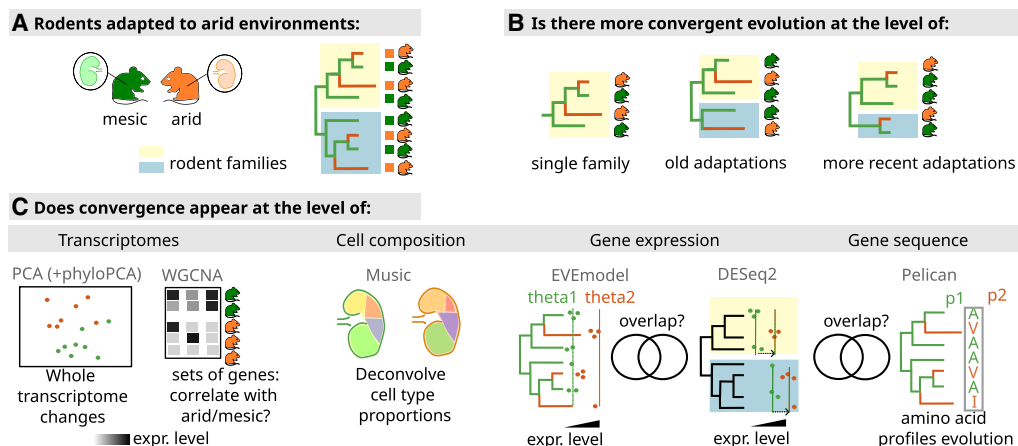


Figure 1. Outline of the analysis of convergent adaptation to arid environment in rodents (A), which aims at quantifying convergence at different levels and phylogenetic scales (B). The methods are indicated in gray, and their main purpose is schematized (C). In particular, the EVE model and Pelican are two phylogenetic methods. EVE models the evolution of expression toward two optima depending on the condition (θ_1 vs. θ_2). Pelican looks for differences in amino acid usage at a given site between two conditions (here, p_1 vs. p_2). DESeq2, when used accounting for family effects, models a similar shift in expression level in each family regardless of the family subtree.

macedonicus locate with mesic species of the genus *Mus*, probably because they are closely related and because *M. macedonicus* lives in the moderately arid mediterranean environment. To investigate the phylogenetic structures in these multivariate comparative data, we then used a phylogenetic principal component analysis (pPCA) (Jombart et al. 2010a,b; Supplemental Methods). We observed that in the Total_expr data set, the third component, which shows a phylogenetic signal, is also correlated with the environment (PC3) (Fig. 3B; Supplemental Fig. S6). In the Murinae_expr data set, the two components with the most variance are correlated with the environment, one with a phylogenetic signal (PC1) but the other in opposition with the phylogenetic signal (PC12) (Fig. 3F; Supplemental Fig. S7). This suggests convergence at the whole-transcriptome level.

By constructing dendrograms based on gene expression distances of DE genes (thus, without phylogenetic correction) we confirm that species adapted to arid environments tend to be closer, but the bootstrap supports are low. Hence, as in the PCA, there is a strong phylogenetic signal mixed with environment signal in expression divergence (Supplemental Figs. S4, S8).

Second, we hypothesized that convergence in gene expression could be detected in functionally related sets of genes that exhibit coordinated changes in expression throughout the phylogeny. We ran correlation network analyses to detect these modules of covarying genes (Horvath 2011). Seven modules in the Total_expr and 10 in the Murinae_expr data set showed a significant correlation with the arid habitat (Fig. 3C,G; Supplemental Table S11). Within each rodent family, the expression level in the Total_expr modules varied consistently with the habitat, confirming that a convergence signal is present alongside the phylogenetic signal (Supplemental Figs. S4, S9). The 10 Murinae_expr modules presented several overrepresented functional categories consistent with the literature mentioned in introduction, such as lipid metabolic processes, cholesterol metabolism, anion transport, and symporter activity (Supplemental Table S10; for top 25 terms, see Supplemental Fig. S10).

Third, we assessed differential expression by pairwise contrasts between arid and mesic conditions. Using the standard tool DESeq2 for differential expression analysis, we accounted for the family as a batch effect to identify genes that were signifi-

cant at the 0.1-adjusted P -value threshold and exhibited a log-fold change greater than 0.4 (for lists and other thresholds, see Supplemental Tables S6, S7). We found 41 genes in the Total_expr data set and 744 genes in the Murinae data set. This is 30.0 and 31.7 times more genes than expected (P -values < 0.0001 estimated from permutations) (Supplemental Fig. S11). Among these genes, we found genes involved in kidney metabolism and kidney adaptation to aridity (e.g., *Slc35b4* and *Slc40a1* in Total_expr; *Aqp2*, *Aqp7*, *Slc14a2*, *Slc8b1*, and *Slc27a2* in Murinae_expr) (Supplemental Material; Supplemental Tables S9, S10). In addition, we applied the EVE phylogenetic modeling method, an approach based on an Ornstein–Uhlenbeck model with two optima that allows statistical assessment of branch-specific shifts in expression while accounting for interspecies variation (Rohlf and Nielsen 2015; Gillard et al. 2021) but works one gene at a time, contrary to DESeq2. After P -value adjustment for multiple testing, we found only two genes in the Total data set and four in the Murinae data set (0.1-adjusted P -value threshold). However, many genes found by DESeq2 are significant with EVE, suggesting that both methods are to some extent congruent (70.1% of DESeq2 significant genes have EVE P -value < 0.05 for Total_expr, 77.6% for Murinae_expr) (Fig. 3D,H; Supplemental Fig. S12A). The remaining genes may reflect different abilities of the two methods to detect phylogenetic structures of convergent expression signals (Fig. 1).

Because the various approaches are consistent with one another, after correcting for the phylogeny, we conclude that the Total_expr rodent data set exhibits a moderate signal for habitat-related convergent evolution of gene expression levels, whereas the Murinae_expr data set displays a stronger signal.

Transcriptome convergence is associated with convergent shifts in kidney cell composition

Tissue bulk RNA-seq data reflect variation in both expression per cell and cell-type composition (abundance). Here, different species may exhibit divergent kidney structures as part of their adaptation to arid environments. We investigated whether tissue composition distinguishes species living in mesic and arid environments by deconvoluting our bulk RNA-seq data with MuSiC (Wang et al.

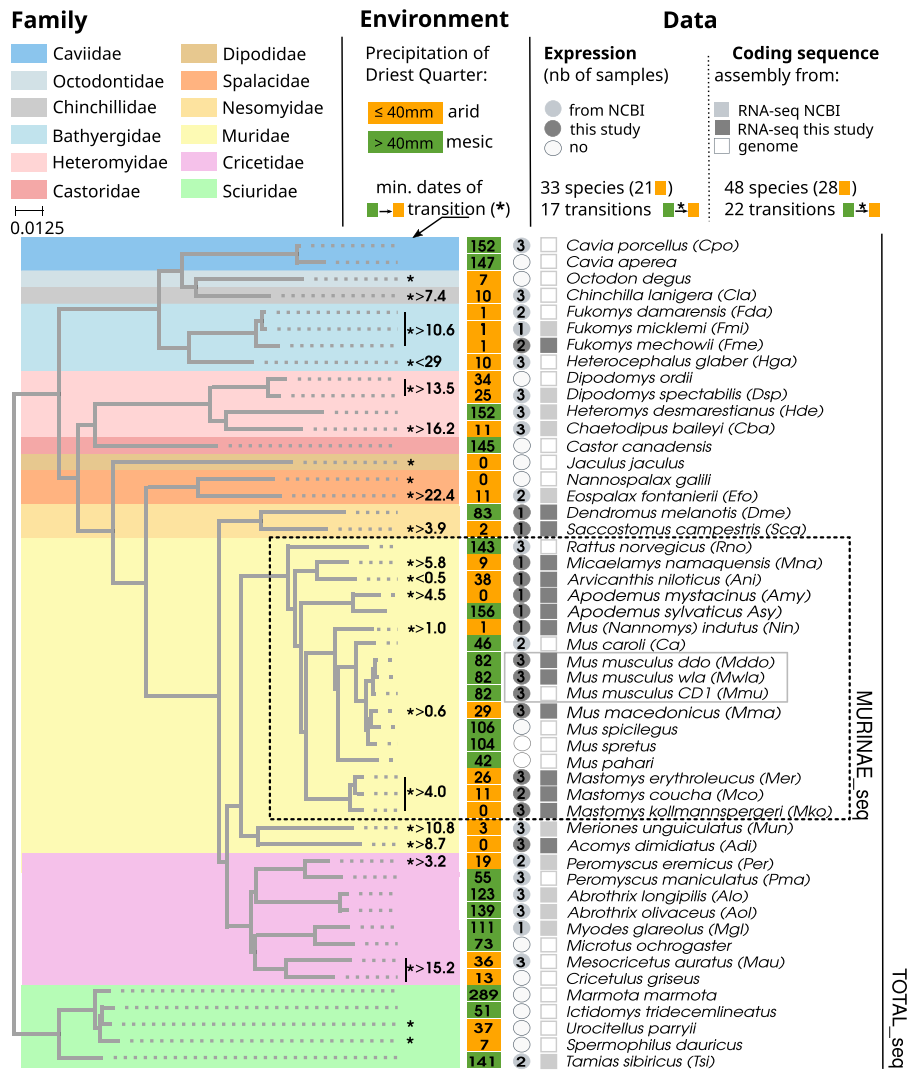


Figure 2. Phylogram showing the species used in the study, made using maximum likelihood on a concatenate of 4065 orthologous gene families. Expression (circles) and coding sequence assemblies (squares) with their source (colors) are indicated. *Mus musculus* strains (gray box, not included in the 33 species count), Murinae_seq data set (dashed lines box), and Total_seq data set (plain line) are delineated. Families and environments are color-coded, and precipitation is indicated. Transitions to arid environments are indicated by stars together with minimum transition ages (or maximum when not available) for data expression analysis.

2019). This method uses cell-type-specific gene expression from single-cell RNA-seq data in a reference data set to characterize cell-type compositions from target bulk RNA-seq data (Fig. 4A). We made several controls to choose the best reference data set (a rat data set from Balzer et al. 2023), and to evaluate the limits of the method (Supplemental Figs. S13–S17; Supplemental Methods). We estimate that the method can be used for moderately distantly related species, because markers of cell types are largely conserved between rats, mice, and hamsters.

In both the Total_expr and the Murinae_expr data set, the first two components of a PCA calculated on these proportions separated species living in arid/mesic environments (Fig. 4B,D; Supplemental Table S12). This discrimination was significant (discriminant analysis [DA], P -values = 0.001 and 0.002 respectively, Monte-Carlo tests based on 1000 replicates). Visually the separation is noticeable but imperfect in the Total data set but is very

clear in the Murinae data set, with the exception of the *M. macedonicus* samples (which resemble mesic species as seen above). Cell types that mostly contributed to this separation were connecting duct, distal convoluted tubule, stroma, intercalated cells, and endothelial cells on the arid side and proximal tubule and loop of Henle cells on the mesic side. Several cell-type-associated P -values were significant after controlling for the family effect (Fig. 4C, linear models). Notably, for certain cell types, we saw the same pattern across most families (e.g., proximal tubule cells decreased in proportion in species living in arid environments) but not for other cell types (e.g., loop of Henle cells tend to increase in proportion in Cricetidae and Heteromyidae but not Muridae) (Supplemental Fig. S18).

Two observations show that this convergent evolution in cell proportions globally impacts the transcriptome. First, variation in cell proportions correlates with coexpression modules: Modules upregulated in arid species correlate positively with cell populations whose proportions are enriched in these species and reversely for modules downregulated (Fig. 4E; Supplemental Fig. S19). Second, marker genes tend to be enriched in differentially expressed genes significantly up- or downregulated in species living in an arid environment, differently depending on the cell population they mark: For instance, intercalated cells marker genes contain upregulated and proximal tubule marker genes downregulated DE genes (Supplemental Fig. S15).

Taken together, these results imply that both Total and Murinae data sets exhibit a significant level of convergent evolution of cell proportions, again with a stronger signal in the Murinae data set. This convergence in cell proportions very likely shapes the global trend of transcriptome convergence pointed above with PCA and gene modules.

Convergent evolution in coding sequences is mosaic

Studies on convergent evolution in coding sequences have shown that convergence on the exact same amino acid occurs but is rare. Few convergent phenotypes, including echolocation or C4 carbon fixation in plants, exhibit convergent substitutions that can be deemed perfect, such that a transition to an identical amino acid is found on all branches where a transition to the new phenotype occurred (Besnard et al. 2009; Marcovitz et al. 2019). Moreover, the exact same change is not needed to accomplish a similar functional effect on a protein. Here, to characterize cases of convergent evolution in coding sequences, we looked for differences in amino acid usage at a given site between arid and mesic species in the

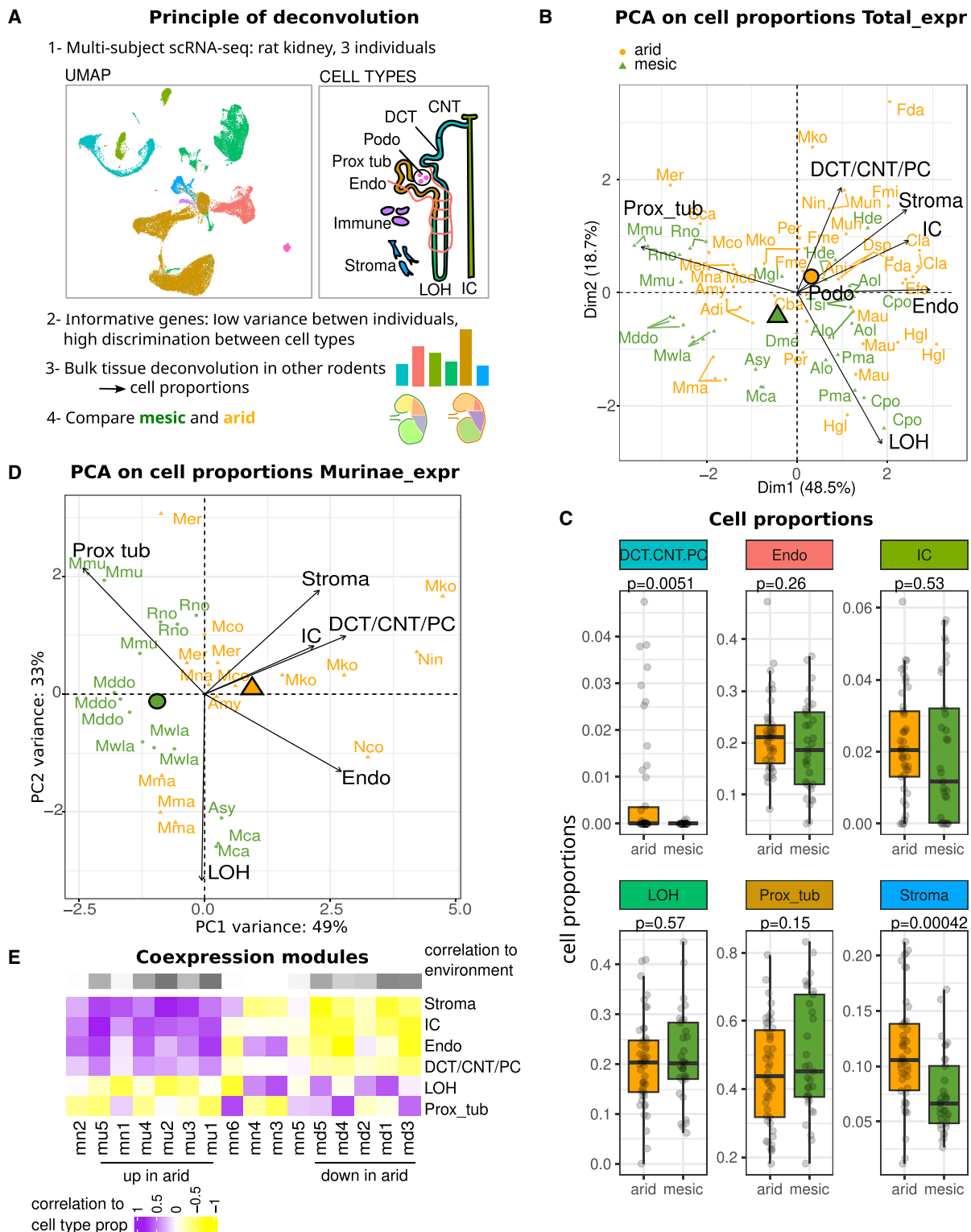


Figure 4. Convergent tissue composition in species living in an arid environment. Cell proportions were estimated by deconvolution using rat scRNA-seq data. (A) Principle of the method. UMAP redrawn from selected samples from Balzer et al. (2023). All cells from the scRNA-seq data except immune cells were used. (B,D) PCA on cell-type proportions is presented for the Total_expr and Murinae_expr data sets. Centroids are represented for species living in a mesic (green) and arid (orange) environment. (C) Cell proportions for species of the Total_expr data set, segregated by habitat. *P*-values of linear models testing for an effect of habitat accounting for the rodent family are indicated. (E) Correlations of Murinae coexpression modules (shown in Fig. 3C) with cell proportions, colored according to key. Cell types are abbreviated as follows (see also schema of panel A): endothelial cells (Endo), stroma cells (Stroma), proximal tubule cells (Prox tub), loop of Henle (LOH), distal convoluted tubule (DCT), connecting tubule (CNT), collecting duct principal cells (PC), and collecting duct intercalated cells (IC), immune cells (Immune).

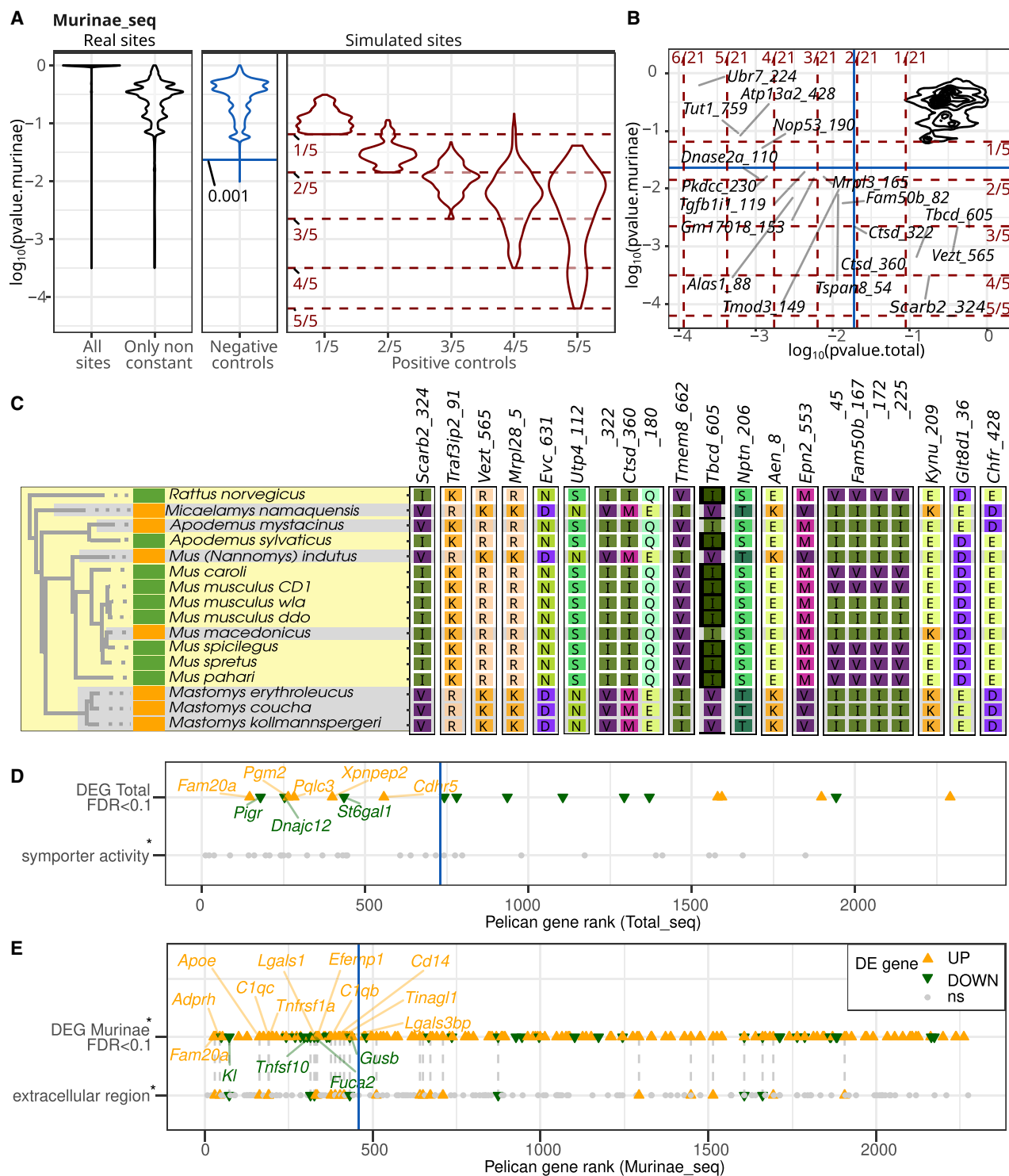


Figure 5. Convergent evolution in coding sequences. (A) Distributions of $\log_{10}(P\text{-value})$ of Pelican tests for real Murinae_seq data (black; all sites and excluding constant sites), simulated negative controls (blue; 1/1000 quantile is indicated), simulated positive controls (red; number of transitions is indicated; dashed lines show the minimum $P\text{-value}$). (B) Scatterplot displaying $\log_{10}(P\text{-value})$ of Pelican tests for all sites common to Murinae_seq (x-axis) and Total_seq (y-axis). The density plot is in black. Blue and dashed red lines as in A for Murinae_seq and as in Supplemental Figure S20A for Total_seq. (C) Top 21 best sites found by Pelican with Murinae_seq data set. (D) Genes ordered according to Pelican's best site-wise $P\text{-value}$ in the Murinae_seq data, with 1/1000 threshold in blue. DE genes indicated by triangles. Genes that belong to selected significant GO-term are indicated. Star indicates a significant enrichment (adjusted $P\text{-value} < 0.1$). (E) Same for Total_seq data.

Altogether, the analysis of expression levels, tissue composition, and genomic convergence suggests a greater level of convergent evolution within a single rodent family, Murinae, than in the total data set. One possible reason is that Murinae have recently diverged and thus share a common genomic background. Another reason may be that their transitions all occurred in a more similar time frame, whereas in the total data set some species belong to lineages that have adapted from mesic ancestors tens of millions of years ago, but others adapted very recently.

Convergent evolution in expression levels is stronger for ancient transitions than for recent transitions

To explore the effect of the age of the transitions, we prepared two data sets of size similar to the Murinae, in terms of number of species and number of transitions to the arid habitat (four transitions), but these transitions are either relatively recent (*recent_expr*, within the same genus and less than 6 MY ago) (Fig. 6A), or more ancient (*ancient_expr*, at the base of a rodent family and/or more than 6 MY ago) (Fig. 6E). In both cases, transitions happening in different rodent families were represented so that the “common background” effect was minimized (three different families for the older transitions, two families for the younger transitions; see later how we controlled for this effect).

These dates were assessed using time-calibrated trees and an extensive review of the literature (Fig. 2; [Supplemental Methods](#); for chronogram, see [Supplemental Fig. S2](#)). It should be noted that some mesic species are present in both data sets, but this should not affect our conclusions, as it is the events happening in the arid branches that interest us. We then applied the same methods as in the previous sections to analyze convergent evolution in expression levels.

In the *recent_expr* data set, we found that the fourth axis of a PCA analysis of expression levels correlated best, albeit imperfectly, with arid/mesic status and explained 7% of the variation (Fig. 6B). In the *ancient_expr* data set, the second axis of the PCA clearly separated arid and mesic species and explained 16% of the variation (Fig. 6F; for main axes in the phylogenetic PCAs correlated with phylogeny, see [Supplemental Figs. S23, S24](#)). This concurs with the number of genes with differential expression found by DESeq2, which was much lower in the *recent_expr* data set than in the *ancient_expr* data set (Fig. 6D,H). The number of genes with convergent evolution found by the EVE model was relatively larger in the *recent_expr* data set, but we show by simulations that this topology tends to favor detection with this method ([Supplemental Methods](#); [Supplemental Fig. S12](#)).

Three gene expression modules were significantly correlated with the arid environment in the *recent_expr* data set, whereas seven were correlated in the *ancient_expr* data set (Fig. 6C,G). They were enriched for relevant functions in the *ancient_expr* but not in the *recent_expr* data set ([Supplemental Fig. S25](#); [Supplemental Table S10](#)). This was also coherent with dendrogram analyses of DE genes ([Supplemental Fig. S26](#)).

Altogether, whole-transcriptome patterns, gene modules, and differential expressed genes detected by both methods show more convergent evolution in ancient transitions than in recent transitions.

Disentangling the contribution of environmental and temporal factors on expression level and cell composition evolution

In prior sections, we emphasized that convergent evolution in gene expression levels is more pronounced within a single family

than throughout the entire data set and that it is also more pronounced for ancient transitions than for more recent transitions. This was noted for data sets that were equivalent in species count and transition count, albeit for a singular instance in each scenario. To generalize this discovery and separate these two effects, we constructed every possible species pair in the data set. Pairs in which both species inherited a change in habitat were eliminated. We quantified the correlation between transcriptomes and between cell compositions for each pair. Of note, we cannot run statistical tests on this analysis because the pairs are not phylogenetically independent.

We observed more variation between kidney transcriptomes of arid species compared with mesic species. This is explained if adaptation to aridity drives kidney diversification and/or drives faster genomic evolution, but mesic species tend to conserve more similar kidneys and/or have slower evolution rates. Whatever its origin, this effect will limit the amount of detectable convergent evolution.

Kidney transcriptomes were more similar for pairs that belonged to the same family than for pairs that did not, as expected from phylogenetic relatedness (Fig. 7A). This effect is even greater for pairs of species independently adapted to the arid environment, with species from different families more dissimilar and species from the same family more similar on average than expected from mesic species. We made a similar observation when comparing cell proportions (Fig. 7B). This is explained if convergent evolution in expression levels and cell proportions is stronger when species evolve independently from ancestors that were more alike.

In order to separate this effect from the age of transition, we split the pairs of species adapted to the arid environment depending on the age of the transition (Fig. 3A). As we saw earlier, transcriptome correlations were higher for pairs of species belonging to the same family but also for more ancient transitions (Fig. 7C). We also observed this when we calculated the correlations only on the union of all the differentially expressed genes found in the previous sections (1325 genes) (Fig. 7D). We observed high correlations for the pairs with ancient transitions, whether the species belong to the same rodent family or not, but observed lower correlations for the pairs with recent transitions. This is explained if convergent evolution in expression levels and cell proportions is stronger between ancient adaptations.

Discussion

We studied repeated evolution during adaptation to aridity, at the macroevolutionary scale, in 12 families of rodents. Our analyses covered transcriptomes, coding sequences and cell-type proportions, across timescales. The strength of our study is the large number of mesic and arid species, including several species sampled in the wild. A caveat in this strategy is that we only have a single individual in several species. Therefore, we cannot study species-specific changes in expression, but we can study convergence at the clade level. Another caveat is that we treated together animals that had different access to water at the time of sampling (from ad libitum drinking in laboratory, to wet and dry season in the wild), which likely reduced our ability to detect convergent changes in expression. Even though some ancestral states may be subject to caution despite our care to annotate them, which may affect the results of phylogenetic methods, we believe the overall message of the analysis is strong because we rely on several other approaches, all of which are congruent.

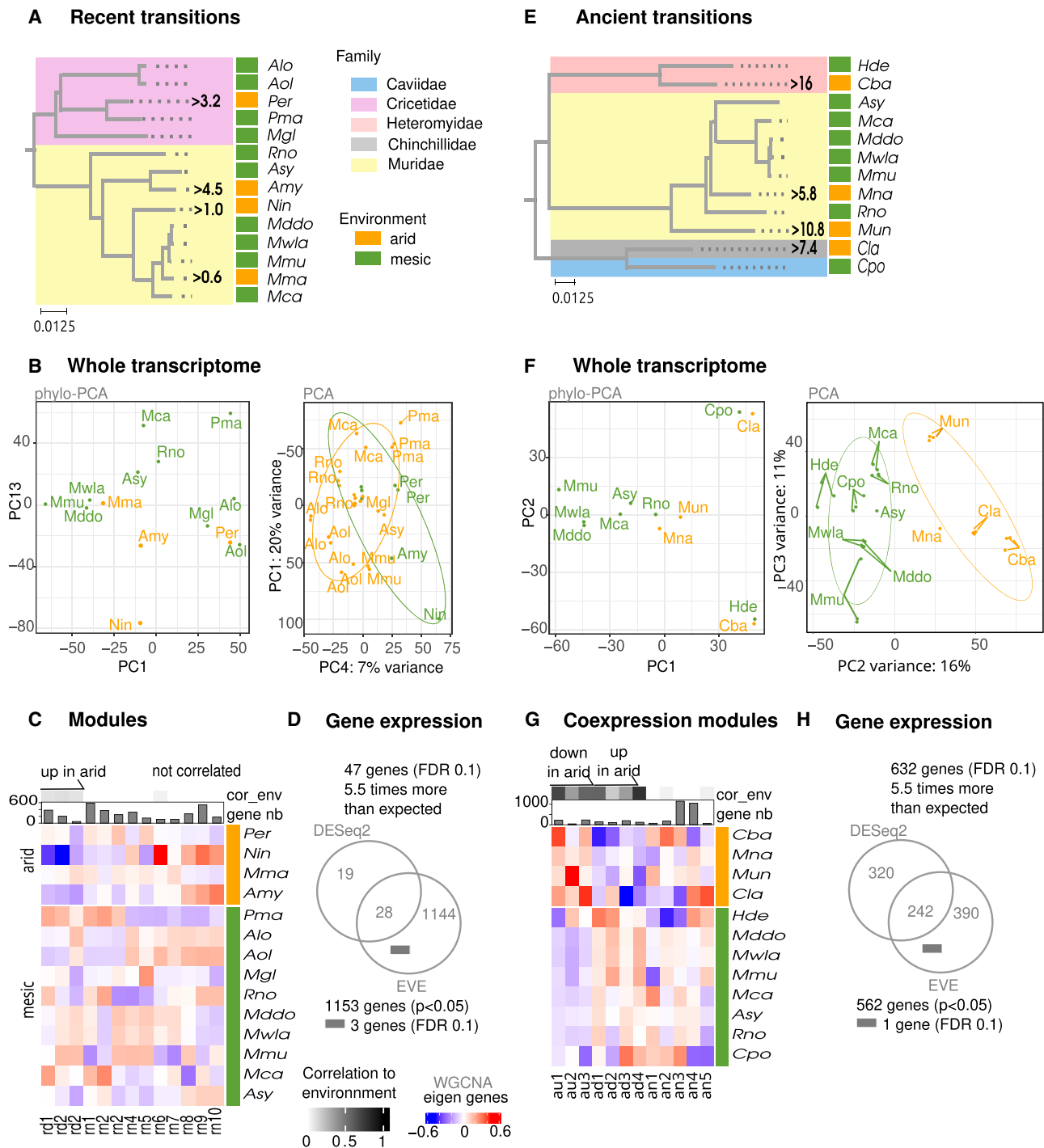


Figure 6. Detection of convergence in data sets with ancient and recent transitions to arid environments. (A) Phylogenetic relationships of the subset with recent (within genera) transitions. The environments and families of the species are color-coded as in Figure 2, and the dates of transitions are indicated. (B) Phylo-PCA plot and PCA plot using family-corrected expression values. (C) Heatmap representing eigengenes per species and per coexpression module. (D) Differentially expressed genes called by DESeq2 and genes detected by the Eve model (genes passing the FDR threshold are only found in the nonoverlapping genes as shown with the small gray bar). (E–H) Same for the set with ancient (between genera) transitions.

A meta-analysis has shown that cases of repeated phenotypic evolution exhibit higher rates of convergent molecular evolution within recently diverged lineages than within lineages that diverged a longer time ago, but the relationship becomes less clear within clades with older divergence (Bohutínská and Peichel

2024). Here we tested the effect of timescales on convergent expression evolution at the macroevolutionary level. We focused on a single organ and a single rodent clade. Compared with the above-mentioned meta-analysis of different works carried out in different clades and for different phenotypic traits, this has the

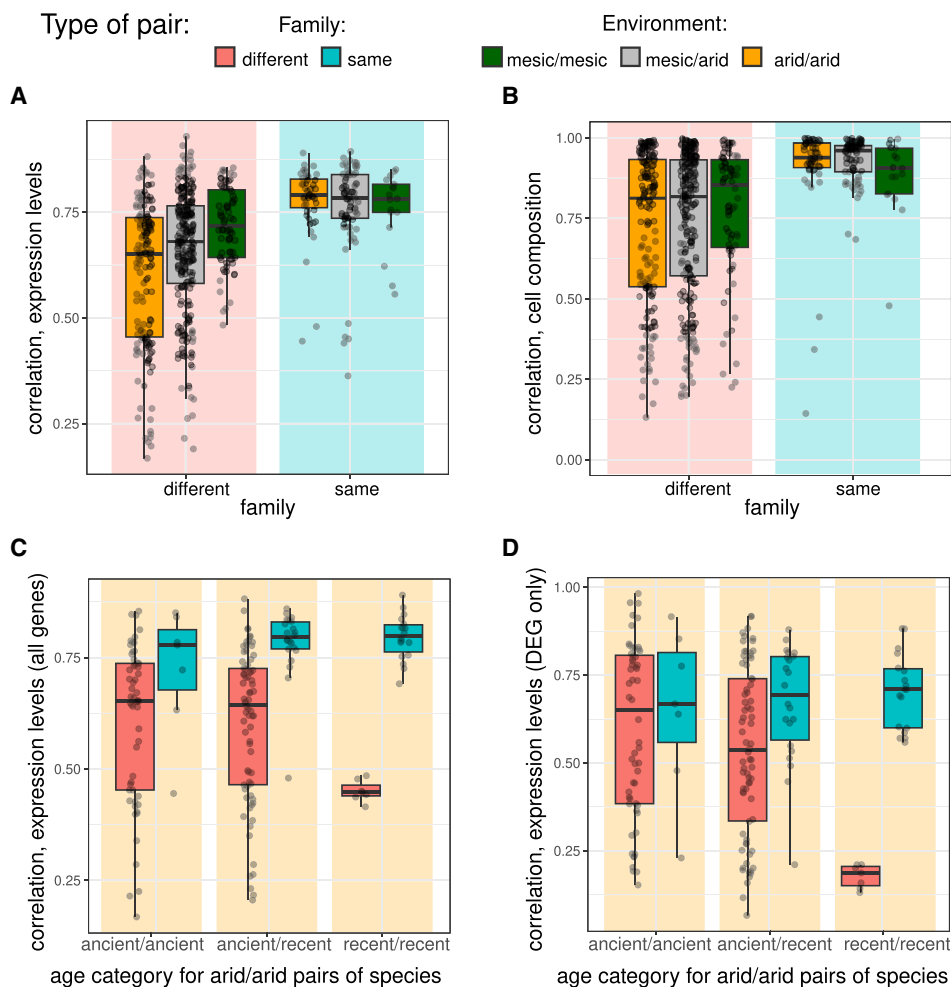


Figure 7. Similarity of expression levels and cell composition in pairs of species. Transcriptome similarities of expression levels (A) and cell-type composition (B) measured with Pearson correlation coefficients for pairs of species living in similar or different environments, belonging to the same or to different rodent families. (C) Transcriptome similarities of expression levels measured for pairs of species living in arid environments, belonging to the same or to different rodent families, after ancient or recent transitions (all independent pairs from Fig. 3A were taken). (D) Same but correlations measured on all differentially expressed genes detected in the study (N = 1325 genes).

advantage of better controlling the confounding effects of differences in polygeny and genome architecture on the level of molecular convergence.

In our type of study focusing on an environmental transition, there are many reasons why convergence may not be found. First, our analysis merged in the same category different ecological realities, associated with different types of challenges. Second, the same environmental challenge may be answered through different phenotypic traits. Third, the same adaptive trait may involve different functional strategies and, therefore, genetic strategies. For example, it is known that different mammals show different structural strategies to improve water retention in the kidney, and different solute carriers have been pointed out in previous studies.

Despite this, we found convergence at all examined biological levels, from bulk kidney transcriptomes down to the coding sequences and expression levels of individual genes. A number of these genes are associated with kidney physiology. Notably, we found in our data several genes and pathways described in recent surveys (Supplemental Material; Rocha et al. 2021; Bittner et al. 2022).

Transcriptome-level signatures reflect convergent changes in cell proportions, suggesting convergent structural adaptations of the kidney

The structure of kidneys varies considerably among mammals (Zhou et al. 2023), with differences in renal histology related to adaptations to arid environments (Bankir and de Rouffignac 1985). Certain differences are species-specific, such as the unique papillary loop of the chinchilla (Chou et al. 1993). Others have been measured across a wide range of species, such as the relative thickness of the medulla, which is proportional to the maximal length of the loop of Henle (Beuchat 1996) and is positively associated with habitat aridity (al-Kahtani et al. 2004). Here, we do observe a signal of convergence in several cell-type proportions, using rat scRNA-seq as a basis for defining cell-type transcriptomes. The proportion of proximal tubule cells (PTs) was larger in mesic species: This was observed independently in different rodent families (Supplemental Fig. S18). Stromal, LOH, and endothelial cells also increase in proportion in arid species, but only in certain families. Together with the diversification and time effect shown in Figure

6, this suggests that adaptation drives strong structural divergence of kidneys in arid species compared with their more monotonic mesic counterparts, but occasionally these structural changes are convergent.

The convergent changes in cell proportions may explain some of the convergent changes in marker genes (Supplemental Material). Because these proportion changes are correlated with coexpression modules, this reiterates the often-overlooked impact of differences in cellular proportions on bulk RNA-seq (Pantalacci and Sémon 2015). To go further, the convergent signal could be refined by more precise cell-type annotation. Renal single-cell RNA sequencing data from multiple rodent species, ideally including arid species, will also be needed to study expression change per cell type versus cell proportion changes. As in the kidney, cellular composition has likely undergone convergent changes in other complex and heterogeneous organs, and the deconvolution approach we present here could help study them.

Convergence in amino acid usage suggests multiple genetic paths repeatedly taken in a mosaic manner

A large proportion of genes showed convergent amino acid substitutions. The enrichment of relevant gene functions like symporter genes or in differentially expressed genes in arid versus mesic species suggests a link with kidney adaptation to aridity. However, we observed that convergence most often does not happen in more than two independent transitions and may affect traits other than the kidney (provided that a gene is expressed at reasonable levels in kidneys, it is in our data set). We expect that the combination of these two effects will strongly diminish the chances to find significant GO terms. Furthermore, we do not rule out that more indirect effects inflate the number of these genes. For example, faster evolutionary rates of habitat-driven speciation may increase the chances to observe convergent epistatic changes in proteins with neutral functional consequences.

Even in the sites with the best scores, we did not observe sites with the exact same amino acid change occurring in all arid species, but changes occurring in four to five clades at best. We performed simulations that refute low sensitivity in our detection. The observation that, for each gene, convergence happens for some clades only indicates that several genetic paths are repeatedly followed but in a mosaic manner, and there is no main path for adaptation, which would be systematically used in all transitions. Consistent with our results, a recent analysis of molecular evolution associated with diverse convergent phenotypes in rodents found very few cases of perfect convergent amino acid evolution (Roycroft et al. 2021).

We were not able to study all the genes in the rodent genome, because we left aside genes that had undergone recent duplications and low-expressed genes whose sequence could not be reconstructed in certain species. It therefore remains possible that examples of perfect convergent amino acid substitutions (or convergent expression) are hidden among the remaining genes.

Effect of temporal factors on the convergent evolution of expression levels

We found a striking coherence among the different methods we used for analyzing expression data sets, even though each has its own unique characteristics. First, with every method, we observed a strong phylogenetic signal in the expression data. Second, similarly to the patterns observed for coding sequence, a mosaic signal of convergence was found comparing gene expression in species

spanning the Rodentia order, whereas more systematic convergence was found focusing on the Murinae family.

We then more directly tested the influence of temporal factors. We observed more convergent changes when we selected species independently adapted from more closely than more distantly related ancestors and when we compared older transitions versus recent transitions. Our study shows that there are many different, yet repeatedly selected, ways to adapt to aridity and that the degree of convergent evolution increases with both the age of the transitions and species relatedness.

The reasons for this excess may differ in the two cases. The convergences that have taken place within the same family are perhaps favored by the fact that these species share a relatively recent common ancestor and therefore share similarities in their mutational landscape, protein interactomes, regulatory pathways, or, even in some cases, alleles. A phylogenetic effect is visible in the coexpression modules, which can be considered as the mark of this common background in gene expression. Our analysis also shows that changes in kidney structural composition increase with time, which will blur the evolution of bulk gene expression levels, reducing our power to detect convergence if no close control species can be included in the study design.

The molecular convergences between distant lineages that have long adapted to aridity could be attributed to the fact that many separate adaptations in various traits have accumulated, increasing the chances of finding some repeatedly. This highlights that the amount of convergent evolution is influenced by both historical contingency, leading to clade-specific adaptations, and timescale.

Methods

Species selection and data preparation

Kidneys sampling

We obtained rodent kidneys of three kinds (Supplemental Tables S1 and S2; for all details, see Supplemental Material). First, some samples came from animals trapped in their natural habitat in six countries. Only one species was trapped on purpose for this study (France). The others were trapped as part of other projects (Senegal, South Africa, Greece, Cameroun, Nigeria, Benin), and kidneys were generously donated to us from collections (Michaux et al 2005; Chevret et al. 2014; Garba et al. 2014; Savassi et al. 2021) or after being preserved for the purpose of this study. Information on the trapping location was available in all instances. Second, other samples came from animals that were part of colonies recently established from wild species (*Acomys dimidiatus*, different *Mus* species, *Fukomys mechowii*). In particular for *Mus* species, this offered the possibility to check if their daily water consumption was consistent with their estimated condition (i.e., mesic or arid). (3) We also added a commercially available mouse strain, as a reference. In all cases, whole kidneys were preserved in RNAlater and stored at -20°C .

Kidney dissection, RNA extraction, and sequencing

To obtain representative expression levels, we removed fat tissue and adrenal gland and used the whole kidney, including for large-sized species (Supplemental Methods). All pieces from a single kidney were lysed in TRIzol with a Precellys homogenizer (Bertin) before precipitation and further purification using the RNeasy mini kit from Qiagen. RNA integrity was controlled on a TapeStation (Agilent Technologies, most RIN was more than 7.8,

occasionally more than 6.5). poly(A)⁺ libraries were prepared with the TruSeq V2 kit (Illumina, nonstranded protocol), starting with 150 ng total RNA before sequencing (Illumina HiSeq 4000, 50 bp single end; one individual per species was resequenced in 100 bp paired-end for transcriptome assembly).

Bioclimatic variables, transitions to arid environments

We obtained the geographical distribution area of each species using GBIF (<https://www.gbif.org/>) data through the *rgbif* package (Chamberlain and Boettiger 2017). Then, for each species, we extracted BIO17 values of its distribution area with the *dismo* package (10.32614/CRAN.package.dismo), which indicates precipitation values of the driest quarter from the international database *worldclim* (<https://www.worldclim.org/data/bioclim.html>). Median values were calculated for each species after excluding samples from zoos, museums, or laboratories. We considered a species as adapted to the arid environment if the median BIO17 is below 40 and to the mesic if the median BIO17 is more than 40.

To obtain the ancestral state reconstruction, we opted for a semiautomated method, based on a broad annotation in many species, which we manually curated for some of our nodes of interest. The annotation of the bioclimatic variables was intersected with a large-scale tree of rodent species (2257 species) (Supplemental Methods; Fabre et al. 2012). We then reconstructed the ancestral states using a maximum likelihood estimation (*ape* R package) (Paradis and Schliep 2019). This provided us with the scaled likelihoods of ancestral states for each internal node of the tree. Based on this extensive automated annotation, we then reviewed individually the transitions based on literature on rodent phylogeny and phylogeography (Supplemental Fig. S1; Supplemental Methods). The minimum and maximum ages of transitions to the arid environment in the expression data set were then estimated from the literature (Supplemental Methods; for 18 articles, see Supplemental Table S5). Transition dates are expected to be older than the node corresponding to the most basal divergence within a family of arid species but younger than the node linking this family with a family of mesic species (Supplemental Fig. S2).

Convergent changes in gene expression data

An automatic workflow was set up using Nextflow (version 19.04.0, April 2019) and is summarized in Supplemental Figure S27.

Published RNA-seq libraries

We interrogated the NCBI for rodent Illumina RNA-seq libraries and selected those with kidney in the metadata. We excluded data from mouse, rat, and pooled species. Whenever possible we selected for each species the three samples with best quality using MultiQC (Ewels et al. 2016). We obtained 60 samples from 21 different species (Supplemental Table S3).

De novo transcriptome assemblies and expression levels

To fairly estimate gene expression levels, we generated de novo transcriptome assemblies for all species, even when genomes were available. We used the selected 60 public samples plus 17 of our samples (Supplemental Table S3) as a basis for an assembly using Trinity version 2.8.5 (Grabherr et al. 2011). We predicted coding sequences from these assemblies and evaluated them with BUSCO version 3.0.2 (Haas et al. 2013; Manni et al. 2021). The quality of the assemblies is summarized in Supplemental Table S13. Expression levels were obtained by mapping the sequence reads against coding sequences from de novo assemblies using kallisto 0.45.1 (Bray et al. 2016) with the default parameters.

Annotation of transcripts

We selected the rodent subset (14 genomes) from EggNOG version 5 (Huerta-Cepas et al. 2019) and used them as a BLASTX database to align the predicted coding sequences (CDSs). For each CDS, we retained the best hit with an *E*-value threshold 1×10^{-6} and assigned it to the corresponding EggNOG cluster. We kept the CDS with the best hit when multiple CDS of the same species were linked to a particular EggNOG cluster.

Preparation of the expression matrices

We isolated 11,437 gene orthogroups with data for at least three species. All transcripts of a given gene were imported using *tximport* package (Soneson et al. 2015) with the option `countsFromAbundance = "lengthScaledTPM"` for additional scaling using the average transcript length to account for gene length differences between species (Supplemental Table S3; Supplemental Data S5). We then performed the following steps on each data set independently.

We first adjusted the expression levels to minimize the influence of phylogenetic effect by applying batch correction using *ComBat_seq* from *sva* package (Zhang et al. 2020). Batch groups are defined based on species phylogenetic relatedness and usually correspond to a rodent family. Because we need at least one arid species and one mesic species in a batch group to perform resampling (see below), we combined two sister families in the same batch when needed (Supplemental Table S3). The correction was applied on all sets except Murinae, in which all species belong to the same family.

Convergence detection by differential and correlation network analyses

On each data set, only genes with nonnull values in all individuals were used, and mitochondrial genes were removed.

We implemented PCA using the *prcomp* function from the *stats* package. Between-class analyses were used to estimate the effect of environmental factors (Dray and Dufour 2007). Dendrograms using expression data were drawn using neighbor-joining on the basis of expression distance matrices (1 – Spearman's correlation coefficient) complemented with 1000 bootstrap analyses (*ape* 5.7) (Paradis and Schliep 2019). We implemented phylogenetic PCA using the *ade4* package (Jombart et al. 2010a,b).

We performed differential expression analyses using the DESeq2 package (Love et al. 2014). Differentially expressed genes were filtered based on \log_2 fold-change threshold of 0.4 corresponding to fold-change greater than 1.3 ("greaterAbs") and adjusted *P*-value < 0.1 (Supplemental Table S7; for results with other thresholds, see Supplemental Table S6). To estimate whether the number of observed DE genes between the arid and mesic species is significantly different from a random observation, we set up a protocol based on permutations inspired by Bittner et al. (2022) that respects the phylogenetic groups (Supplemental Table S3; Supplemental Methods).

We modeled the evolution of expression using the *EVE* model R package (Gillard et al. 2021) using norm-transformed counts of expression (method *ntd* from DESeq2 package). We performed the "twoThetaTest" method contrasting one homogeneous tendency on all branches with two phenotype-specific tendencies (mesic vs. arid) (Supplemental Table S8). We performed simulations to study the sensitivity of the model for different tree topologies (Supplemental Methods).

We searched for modules of genes with correlated expression values with a weighted gene coexpression network analysis (WGCNA version 1.72) (Horvath 2011). We used normalized counts obtained with DESeq2 and the top 50% most variable genes

based on *varianceStabilizingTransformation* from DESeq2. We selected soft-thresholding power from five to 20 based on the *pickSoftThreshold* function, and we used the “signed” network and a *minModuleSize*=30 in the *blockwiseModules* function in WGCNA.

We computed functional enrichment for Gene Ontology terms using *EnrichGO* function from the ClusterProfiler package (Wu et al. 2021) on lists of differentially expressed genes and modules of coexpressed genes significantly correlated with arid/mesic habitat. We developed an in-house script to obtain networks of gene enrichment results visually similar to (Pollard et al. 2024). A network displays significantly overrepresented GO terms (P adj < 0.1) as nodes, with edge weights reflecting shared gene percentages used for clustering by the Louvain community detection algorithm.

We established a list of renal marker genes with cell type and renal segments based on a search of databases and literature as detailed in the Supplemental Methods (Supplemental Appendix 2; Supplemental Table S9).

Deconvolutions

To determine whether changes in cell proportions occurred between mesic and arid species, we tested several computational methods to infer cell-type proportions from bulk transcriptomic data, based on marker genes and on species with kidney scRNA-seq currently available in rodents (mouse, rat, and hamster) (Supplemental Methods). Our benchmark shows that MuSiC (Wang et al. 2019) applied with scRNA-seq data taken from rat kidney (Balzer et al. 2023) was best (Supplemental Methods). These data contain 48,744 cells for three lean and untreated individuals (obtained from the NCBI Gene Expression Omnibus [GEO; <https://www.ncbi.nlm.nih.gov/geo/>] under accession number GSE209821), with which we redid the UMAP using Seurat package (Hao et al. 2021) and transferred the annotations from the publication before estimating cell proportions.

Detection of convergent changes in protein sequences

Coding sequences from genome assemblies

We retrieved coding sequences from 24 published rodent genomes (Ensembl release 99) (Harrison et al. 2024). For eight species, genome and kidney transcriptome assemblies were available (Supplemental Table S1). In these cases, cDNA Ensembl sequences were favored over transcriptomes in the sequence analysis. We assigned these coding sequences to EggNOG groups as described above (see Annotation of Transcripts).

Sequence alignments

We grouped coding sequences associated with the same EggNOG group into gene families. We removed families with fewer than three species and sequences within families if their length was <70% of the median length of the family. We aligned the remaining protein sequences with MAFFT (version 7.313, with the options `--localpair --maxiterate 1000`) (Katoh and Standley 2013) and cleaned the alignments with HmmCleaner (version 0.180750) (Di Franco et al. 2019). We discarded sequences for which >50% of the positions were removed by HmmCleaner and amino acid sites with >10% gaps. We obtained multiple alignments for 11,437 sets of orthologs, ranging from three to 48 species (plus two strains) (Supplemental Data S6).

Phylogenetic reconstruction

With a custom script, we back-translated the amino acid alignments of the 4065 complete gene families (with 50 sequences)

and retained only the sites without gaps for phylogenetic analysis (Supplemental Methods). Ten subsets were extracted from these families. For each subset, we chose randomly 500 genes and 200 sites per gene and then concatenated the 100,000 sites (using *catfasta2phym.pl*; <https://github.com/nylander/catfasta2phym>). For genes shorter than 200 sites, all sites were retained in the concatenate. Phylogenetic reconstruction was performed using RAXML-NG software (Kozlov et al. 2019) with the options `--all` and `--model LG+G`. We obtained 10 trees for which we estimated the likelihood of the complete data set. We retained the best tree for Figure 2 and for detecting convergent sequence evolution. The corresponding phylogeny built from our assemblies is congruent with a large-scale rodent phylogeny (Supplemental Fig. S2; for a cophylogeny done with apes, see Paradis and Schliep [2019]).

Detection of convergent sites

We used Pelican (Duchemin et al. 2023) to detect convergent changes in cleaned protein alignments using “multinomial-filter = 0.8” (Supplemental Data S7; Supplemental Methods). For each data set, we filtered the gene families based on the total numbers of mesic and arid species (see Results). We simulated positive controls of convergence with a custom script and negative controls with the Pastek simulator (for details, see Supplemental Methods; <https://gitlab.in2p3.fr/pveber/pastek>, commit: d267be5f). This allowed us to define a threshold of detection and to obtain genes for which one site at least passed the threshold. We performed over-representation analysis using the Gene Ontology database (ClusterProfiler) (Wu et al. 2021).

Protection of human subjects and animals in research

This study used collections of samples as well as specimens sampled in the frame of this study. This was performed in strict respect of legislation, regarding animal trapping, animal care, animal killing, sample transport, and respect of the Nagoya protocol. Specific accreditations are listed below. More details can be found in Supplemental Material (Supplemental Appendix 1).

Collections

Collection in Creta was performed by NHMC under the provisions of the Presidential Decree 67/81.

For collections from the Republic of Benin, access and sharing of advantages was agreed by the government of the Republic of Benin (file 608/DGEFC/DCPRN/PF-APA/SA).

Benin, Cameroon, and Senegal samples were part of the “CBGP–small mammal collection” (<https://doi.org/10.15454/WWNUP0>).

Sampling in the wild

None of the rodent species is protected, and every animal-related procedure was performed according to official ethical guidelines (Sikes and Animal Care and Use Committee of the American Society of Mammalogists 2016).

South African rodent specimens were sampled under permit number JM 1193/2017, issued by the Free State Department of Economic, Small Business Development, Tourism and Environmental Affairs (DESTEA) in Bloemfontein (Free State, South Africa) and exported to France under permit JM 3007/2017, also issued by DESTEA.

Sampling in Senegal were conducted following official regulations (Centre de Biologie pour la Gestion des Populations (CBGP): Agrément pour l'utilisation d'animaux à des fins scientifiques D-34-169-003) of the relevant institutional committee (Regional

Head of the Veterinary Service, Hérault, France). All transfer and conservation procedures were performed in accordance with current Senegalese and French legislation (by the Senegalese “Direction Nationale des Eaux et Forêts, Chasses Et Conservation Des Sols,” partner of the “CERISE” project, which funded this trapping campaign (AAPSCEN-20B III) and the French “Direction Départementale de la Protection des Populations, Hérault” for sample importation authorization.

Sampling in laboratories

Animals were maintained and sacrificed in strict accordance with the European guidelines 2010/63/UE and the local legislation, including in breeding species issued from wild life (France, accreditation SPE-2014-001 69-148; Czech Republic, accreditation 22395/2014-MZE-17214).

Data access

All raw sequencing data generated in this study have been submitted to the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena/browser/>) under accession number PRJEB54931. Previously published cDNA libraries and expression raw data are listed with accession numbers in Supplemental Table S3. All codes are available in a GitLab repository (https://gitbio.ens-lyon.fr/LBMC/cigogne/convergent_aridity_2024) and in the Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank the following for collecting and/or giving access to their live collection: Pascale Chevret (LBBE, France), Frédéric Delsuc, Nico Avenant, and Lionel Hautier (ISEM, France) for Southern Africa samples; Gauthier Dobigny (IRD-CBGP, France), Philippe Gauthier (IRD-CBGP), and Caroline Tataru (INRAE-CBGP) for Cameroon and Benin samples; Radim Sumbera and Lucie Plestilova for the *Fukomys* species; Petros Lymberakis (Natural History Museum of Crete, University of Crete) for *Apodemus mystacinus*; Laurent Granjon, Madougou Garba, Youssou Niang, Mamadou Kane, Aliou Sow, Moussa Sall (IRD-CBGP), Cheikh Niang (UGB), Kodé Fall (IFAN-UCAD), and the CERISE project (AAPSCEN-20B III) for Senegal samples; and François Bonhomme for ISEM collections. The Benin, Cameroon, and Senegal samples are part of the “CBGP–small mammal collection” (<https://doi.org/10.15454/WWNUPO>). We acknowledge support from the Centre Blaise Pascal de Simulation et de Modélisation Numérique (CBPSMN) of the ENS de Lyon and the French Institute of Bioinformatics (IFB CNRS UMS 3601) for computing resources, as well as the GenomeEast IGBMC sequencing platform. This project was supported by an Agence Nationale de la Recherche (ANR) grant to B.B., M.S., and S.P. (ANR-15-CE32-0005). We thank P. Veber and L. Duchemin for fruitful discussions and the reviewers whose insightful comments helped to improve the manuscript.

Author contributions: S.P., D.C., and P.C. designed, realized, and coordinated species sampling; S.P. and M.M. carried out sample preparation; D.C., C.R., and J.G. processed the data; D.C., C.R., J.B., L.G., B.B., and M.S. analyzed the data; D.C., C.R., L.G., B.B., S.P., and M.S. interpreted the results and wrote and edited the manuscript; and S.P. and M.S. supervised the study. All authors read, corrected, and approved the final manuscript.

References

- Aghová T, Kimura Y, Bryja J, Dobigny G, Granjon L, Kergoat GJ. 2018. Fossils know it best: using a new set of fossil calibrations to improve the temporal phylogenetic framework of murid rodents (Rodentia: Muridae). *Mol Phylogenet Evol* **128**: 98–111. doi:10.1016/j.ympev.2018.07.017
- al-Kahtani MA, Zuleta C, Caviedes-Vidal E, Garland T Jr. 2004. Kidney mass and relative medullary thickness of rodents in relation to habitat, body size, and phylogeny. *Physiol Biochem Zool* **77**: 346–365. doi:10.1086/420941
- Balzer MS, Pavkovic M, Frederick J, Abedini A, Freyberger A, Vienenkötter J, Mathar I, Siudak K, Eitner F, Sandner P, et al. 2023. Treatment effects of soluble guanylate cyclase modulation on diabetic kidney disease at single-cell resolution. *Cell Rep Med* **4**: 100992. doi:10.1016/j.xcrm.2023.100992
- Bankir L, de Rouffignac C. 1985. Urinary concentrating ability: insights from comparative anatomy. *Am J Physiol* **249**: R643–R666. doi:10.1152/ajpregu.1985.249.6.R643
- Besnard G, Muasya AM, Russier F, Roalson EH, Salamin N, Christin P-A. 2009. Phylogenomics of C₄ photosynthesis in sedges (Cyperaceae): multiple appearances and genetic convergence. *Mol Biol Evol* **26**: 1909–1919. doi:10.1093/molbev/msp103
- Beuchat CA. 1996. Structure and concentrating ability of the mammalian kidney: correlations with habitat. *Am J Physiol* **271**: R157–R179. doi:10.1152/ajpregu.1996.271.1.R157
- Bittner NKJ, Mack KL, Nachman MW. 2021. Gene expression plasticity and desert adaptation in house mice. *Evolution (N Y)* **75**: 1477–1491. doi:10.1111/evo.14172
- Bittner NKJ, Mack KL, Nachman MW. 2022. Shared patterns of gene expression and protein evolution associated with adaptation to desert environments in rodents. *Genome Biol Evol* **14**: evac155. doi:10.1093/gbe/evac155
- Blumstein DM, MacManes MD. 2023. When the tap runs dry: the physiological effects of acute experimental dehydration in *Peromyscus eremicus*. *J Exp Biol* **226**: jeb246386. doi:10.1242/jeb.246386
- Bohutínská M, Peichel CL. 2024. Divergence time shapes gene reuse during repeated adaptation. *Trends Ecol Evol (Amst)* **39**: 396–407. doi:10.1016/j.tree.2023.11.007
- Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525–527. doi:10.1038/nbt.3519
- Brown AP, McGowan KL, Schwarzkopf EJ, Greenway R, Rodriguez LA, Tobler M, Kelley JL. 2019. Local ancestry analysis reveals genomic convergence in extremophile fishes. *Philos Trans R Soc Lond B Biol Sci* **374**: 20180240. doi:10.1098/rstb.2018.0240
- Chamberlain SA, Boettiger C. 2017. R Python, and Ruby clients for GBIF species occurrence data. *PeerJ* **5**: e3304v1. doi:10.7287/peerj.preprints.3304v1
- Chaturvedi S, Gompert Z, Feder JL, Osborne OG, Muschick M, Riesch R, Soria-Carrasco V, Nosil P. 2022. Climatic similarity and genomic background shape the extent of parallel adaptation in *Timema* stick insects. *Nat Ecol Evol* **6**: 1952–1964. doi:10.1038/s41559-022-01909-6
- Cheng J, Peng X, Li H, Feijó A, Xia L, Shenbrot GI, Ge D, Wen Z, Wang D, Yang Q. 2023. Similar adaptive mechanism but divergent demographic history of four sympatric desert rodents in Eurasian inland. *Commun Biol* **6**: 33. doi:10.1038/s42003-023-04415-y
- Chevret P, Robinson TJ, Perez J, Veyrunes F, Britton-Davidian J. 2014. A phylogeographic survey of the pygmy mouse *Mus minutoides* in South Africa: taxonomic and karyotypic inference from cytochrome b sequences of museum specimens. *PLoS One* **9**: e98499. doi:10.1371/journal.pone.0098499
- Chou CL, Nielsen S, Knepper MA. 1993. Structural-functional correlation in chinchilla long loop of Henle thin limbs: a novel papillary subsegment. *Am J Physiol* **265**: F863–F874. doi:10.1152/ajprenal.1993.265.6.F863
- Corral-Lopez A, Bloch NI, van der Bijl W, Cortazar-Chinarro M, Szorkovszky A, Kotrschal A, Darolti I, Buechel SD, Romenskyy M, Kolm N, et al. 2024. Functional convergence of genomic and transcriptomic architecture underlies schooling behaviour in a live-bearing fish. *Nat Ecol Evol* **8**: 98–110. doi:10.1038/s41559-023-02249-9
- Di Franco A, Poujol R, Baurain D, Philippe H. 2019. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol Biol* **19**: 21. doi:10.1186/s12862-019-1350-2
- Dray S, Dufour A-B. 2007. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* **22**: 1–20. doi:10.18637/jss.v022.i04
- Duchemin L, Lanore V, Veber P, Boussau B. 2023. Evaluation of methods to detect shifts in directional selection at the genome scale. *Mol Biol Evol* **40**: msac247. doi:10.1093/molbev/msac247

- Ewels P, Magnusson M, Lundin S, Källér M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**: 3047–3048. doi:10.1093/bioinformatics/btw354
- Fabre P-H, Hautier L, Dimitrov D, Douzere EJP. 2012. A glimpse on the pattern of rodent diversification: a phylogenetic approach. *BMC Evol Biol* **12**: 88. doi:10.1186/1471-2148-12-88
- Garba M, Dalecky A, Kadaoure I, Kane M, Hima K, Veran S, Gagare S, Gauthier P, Tatar C, Rossi J-P, et al. 2014. Spatial segregation between invasive and native commensal rodents in an urban environment: a case study in Niamey, Niger. *PLoS One* **9**: e110666. doi:10.1371/journal.pone.0110666
- Gillard GB, Grønvoold L, Røsæg LL, Holen MM, Monsen Ø, Koop BF, Rondeau EB, Gundappa MK, Mendoza J, Macqueen DJ, et al. 2021. Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biol* **22**: 103. doi:10.1186/s13059-021-02323-0
- Giorello FM, Feijoo M, D'Elía G, Naya DE, Valdez L, Opazo JC, Lessa EP. 2018. An association between differential expression and genetic divergence in the Patagonian olive mouse (*Abrothrix olivacea*). *Mol Ecol* **27**: 3274–3286. doi:10.1111/mec.14778
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol* **29**: 644–652. doi:10.1038/nbt.1883
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**: 1494–1512. doi:10.1038/nprot.2013.084
- Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell* **184**: 3573–3587. e29. doi:10.1016/j.cell.2021.04.048
- Harrison PW, Amode MR, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, Becker A, Bennett R, Berry A, Bhairi J, et al. 2024. Ensembl 2024. *Nucleic Acids Res* **52**: D891–D899. doi:10.1093/nar/gkad1049
- Horvath S. 2011. *Weighted network analysis: applications in genomics and systems biology*. Springer Science & Business Media, New York.
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**: D309–D314. doi:10.1093/nar/gky1085
- Jombart T, Balloux F, Dray S. 2010a. *Adephylo*: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* **26**: 1907–1909. doi:10.1093/bioinformatics/btq292
- Jombart T, Pavoine S, Devillard S, Pontier D. 2010b. Putting phylogeny into the analysis of biological traits: a methodological approach. *J Theor Biol* **264**: 693–701. doi:10.1016/j.jtbi.2010.03.038
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780. doi:10.1093/molbev/mst010
- Kim C-S, Shin D-M. 2016. Improper hydration induces global gene expression changes associated with renal development in infant mice. *Genes Nutr* **11**: 28. doi:10.1186/s12263-016-0544-0
- Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455. doi:10.1093/bioinformatics/btz305
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8
- MacManes MD, Eisen MB. 2014. Characterization of the transcriptome, nucleotide sequence polymorphism, and natural selection in the desert adapted mouse *Peromyscus eremicus*. *PeerJ* **2**: e642. doi:10.7717/peerj.642
- Manni M, Berkeley MR, Seppely M, Zdobnov EM. 2021. BUSCO: assessing genomic data quality and beyond. *Curr Protoc* **1**: e323. doi:10.1002/cpz1.323
- Marcovitz A, Turakhia Y, Chen HI, Gloude-mans M, Braun BA, Wang H, Bejerano G. 2019. A functional enrichment test for molecular convergent evolution finds a clear protein-coding signal in echolocating bats and whales. *Proc Natl Acad Sci* **116**: 21094–21103. doi:10.1073/pnas.1818532116
- Marra NJ, Romero A, DeWoody JA. 2014. Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed by RNA-seq. *Mol Ecol* **23**: 2699–2711. doi:10.1111/mec.12764
- Michaux J, Bellinvia E, Lymberakis P. 2005. Taxonomy, evolutionary history and biogeography of the broad-toothed field mouse (*Apodemus mystacinus*) in the eastern Mediterranean area based on mitochondrial and nuclear genes. *Biol J Linn Soc* **85**: 53–63. doi:10.1111/j.1095-8312.2005.00469.x
- Pantalacci S, Sémon M. 2015. Transcriptomics of developing embryos and organs: a raising tool for evo-devo. *J Exp Zool B Mol Dev Evol* **324**: 363–371. doi:10.1002/jez.b.22595
- Pantalacci S, Guéguen L, Petit C, Lambert A, Peterková R, Sémon M. 2017. Transcriptomic signatures shaped by cell proportions shed light on comparative developmental biology. *Genome Biol* **18**: 29. doi:10.1186/s13059-017-1157-7
- Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528. doi:10.1093/bioinformatics/bty633
- Peng X, Cheng J, Li H, Feijó A, Xia L, Ge D, Wen Z, Yang Q. 2023. Whole-genome sequencing reveals adaptations of hairy-footed jerboas (*Dipus, Dipodidae*) to diverse desert environments. *BMC Biol* **21**: 182. doi:10.1186/s12915-023-01680-5
- Pollard MD, Meyer WK, Puckett EE. 2024. Convergent relaxation of molecular constraint in herbivores reveals the changing role of liver and kidney functions across mammalian diets. *Genome Res* **34**: 2176–2189. doi:10.1101/gr.278930.124
- Rocha JL, Godinho R, Brito JC, Nielsen R. 2021. Life in deserts: the genetic basis of mammalian desert adaptation. *Trends Ecol Evol* **36**: 637–650. doi:10.1016/j.tree.2021.03.007
- Rocha J, Silva P, Santos N, Nakamura M, Afonso S, Qniba A, Boratynski Z, Sudmant PH, Brito JC, Nielsen R, et al. 2023. North African fox genomes show signatures of repeated introgression and adaptation to life in deserts. *Nat Ecol Evol* **7**: 1267–1286. doi:10.1038/s41559-023-02094-w
- Rohlf RV, Nielsen R. 2015. Phylogenetic ANOVA: the expression variance and evolution model for quantitative trait evolution. *Syst Biol* **64**: 695–708. doi:10.1093/sysbio/syv042
- Roycroft E, Achmadi A, Callahan CM, Esselstyn JA, Good JM, Moussalli A, Rowe KC. 2021. Molecular evolution of ecological specialisation: genomic insights from the diversification of murine rodents. *Genome Biol Evol* **13**: evab103. doi:10.1093/gbe/evab103
- Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**: 74–78. doi:10.1126/science.aat7244
- Savassi BAES, Dobigny G, Etougbéché JR, Avocegan TT, Quinsou FT, Gauthier P, Ibiokounlé M, Moné H, Mouahid G. 2021. *Mastomys natalensis* (Smith, 1834) as a natural host for *Schistosoma haematobium* (Bilharz, 1852) Weinland, 1858 x *Schistosoma bovis* Soinso, 1876 introgressive hybrids. *Parasitol Res* **120**: 1755–1770. doi:10.1007/s00436-021-07099-7
- Schwimmer H, Haim A. 2009. Physiological adaptations of small mammals to desert ecosystems. *Integr Zool* **4**: 357–366. doi:10.1111/j.1749-4877.2009.00176.x
- Sikes RS, Animal Care and Use Committee of the American Society of Mammalogists. 2016. 2016 guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *J Mammal* **97**: 663–688. doi:10.1093/jmammal/gyw078
- Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**: 1521. doi:10.12688/f1000research.7563.1
- Tigano A, Colella JP, MacManes MD. 2020. Comparative and population genomics approaches reveal the basis of adaptation to deserts in a small rodent. *Mol Ecol* **29**: 1300–1314. doi:10.1111/mec.15401
- Wang X, Park J, Susztak K, Zhang NR, Li M. 2019. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* **10**: 380. doi:10.1038/s41467-018-08023-x
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**: 100141. doi:10.1016/j.xinn.2021.100141
- Yuan S, Zhang R, Jin Y, Li X, Li L, Zhang D, Ling Y, Zhang K, Wu X, Zhang X, et al. 2025. Multiomic analyses reveal the molecular mechanisms of arid adaptation in a desert rodent species. *Mol Biol Evol* **42**: msaf230. doi:10.1093/molbev/msaf230
- Zancolli G, Reijnders M, Waterhouse RM, Robinson-Rechavi M. 2022. Convergent evolution of venom gland transcriptomes across Metazoa. *Proc Natl Acad Sci* **119**: e2111392119. doi:10.1073/pnas.2111392119
- Zhang Y, Parmigiani G, Johnson WE. 2020. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* **2**: lqaa078. doi:10.1093/margab/lqaa078
- Zhou X, Rong W, Guo B, He X, Cao L, Zheng Y, Xu S, Yang G, Ren W. 2023. The evolution of the discrete multiretulate kidney in mammals from ecological and molecular perspectives. *Genome Biol Evol* **15**: evad075. doi:10.1093/gbe/evad075

Received October 2, 2024; accepted in revised form January 15, 2026.