Title: Convergent genomic trajectories shape adaptation to life on land across animal lineages

Authors: Gemma I. Martínez-Redondo[1,2] †*, Klara Eleftheriadi[1]†*, Judit Salces-Ortiz[1], Nuria Escudero[1], Fernando Ángel Fernández-Álvarez[3,4], Belén Carbonetto[1], Carlos Vargas-Chávez[1], Raquel García-Vernet[1,2], Javier Palma-Guerrero[1,5], Libe Rentería[1], Iñaki Rojo[6], Cristina Chiva[7,8], Eduard Sabidó[7,8], Aureliano Bombarely[9] & Rosa Fernández[1]*

Affiliations:

[1] Metazoa Phylogenomics and Genome Evolution Lab, Institute of Evolutionary Biology (CSIC-UPF), Barcelona (Spain)

[2] Universitat de Barcelona, Barcelona (Spain)

[3] Centro Oceanográfico de Gijón, Instituto Español de Oceanografía, Gijón (Spain)

[4] Institut de Ciències del Mar (ICM-CSIC), Barcelona (Spain)

[5] Current address: Research Institute of Organic Agriculture (FiBL), Aargau (Switzerland)

[6] Citizen scientist, Sopelana (Spain)

[7] Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona (Spain)

[8] Universitat Pompeu Fabra (UPF), Barcelona (Spain)

[9] Instituto de Biología Molecular y Celular de Plantas, Valencia (Spain)

†These authors contributed equally to this work, order was chosen by dice rolling.

* Corresponding authors: gemma.martinez@ibe.upf-csic.es, klara.eleftheriadi@ibe.upf-csic.es, rosa.fernandez@ibe.upf-csic.es

Abstract: How animals repeatedly adapted to life on land is a central question in evolutionary biology. While terrestrialisation occurred independently across animal phyla, it remains unclear whether shared genomic mechanisms underlie these transitions. Here, we combined large-scale comparative genomics, machine learning, and multi-omics data, including proteomics and transcriptomics from stress experiments relevant to terrestrial environmental challenges in 17 species, to investigate the genomic basis of animal terrestrial adaptation. Gene co-expression networks revealed that genes relevant to stress were largely lineage-specific, yet converged in function through the co-option of gene families pre-dating terrestrialisation events. Phylogenomic and machine learning analyses supported a dominant role for early-evolving genes, enriched in stress-related functions, paired with a higher gene loss than gain at terrestrialisation nodes. Our findings support a model of lineage-specific genomic changes involving mostly conserved genes that converged at the functional level during the independent transitions to terrestrial life.

Main Text

The transition of animals to land (terrestrialisation) is one of the most transformative events in evolutionary history. Coming from relatively stable marine environments, animals underwent radical physiological and morphological adaptations to cope with the environmental challenges of life on land, including desiccation, higher oxygen concentration, ultraviolet radiation, and changes in osmotic pressure and chemical exposure[1–5]. As such, terrestrialisation represents not just a change in habitat, but a major functional reconfiguration of animal biology. Fossil evidence suggests the first land exploration by marine arthropod relatives occurred around 470 million years ago[4–6], likely driven by the need to mate or lay eggs out of predators' reach. Over millions of years, terrestrial habitats became progressively inhabited as environments became more hospitable due to the biota's own influence[1,7,8]. Unlike plants (*sensu stricto*), which terrestrialised only a few times, animals (Metazoa) have independently transitioned to land several times across at least nine out of the 31 animal phyla (Fig. 1a,b). Terrestrialisation in animals is highly heterogeneous, with many species remaining semi-terrestrial, requiring water at specific life stages, while only a few complete their entire life cycles on land[10,11]. These transitions occurred via direct marine-to-terrestrial routes or freshwater intermediaries utilised by certain groups like platyhelminthes, molluscs, crabs, and vertebrates[12–15] (Fig. 1c). This rich tapestry of terrestrialisation, characterised by diverse dependencies on water, varied environmental routes to land, and staggered timelines, underscores the multifaceted nature of

this evolutionary achievement.

Most previous studies have focused on individual species[16,17], specific phyla[18–21] or particular gene families[22,23], without assessing whether genomic adaptation to terrestrial life involved convergent genomic changes across all animal phyla, or arose through lineage-specific mechanisms. To address this, we conducted a broad comparative analysis of nearly 1,000 animal species, integrating genome-wide data with expression profiles from experiments that mimic environmental stressors faced during terrestrialisation[24–27,28,29], along with proteomic data, in diverse aquatic and terrestrial invertebrates. We also identified gene families associated with habitat type through machine learning. By tracing the evolutionary origins and functions of these genes, we provide a comprehensive view of how animals have genomically adapted to life on land through repeated, but distinct, evolutionary routes.

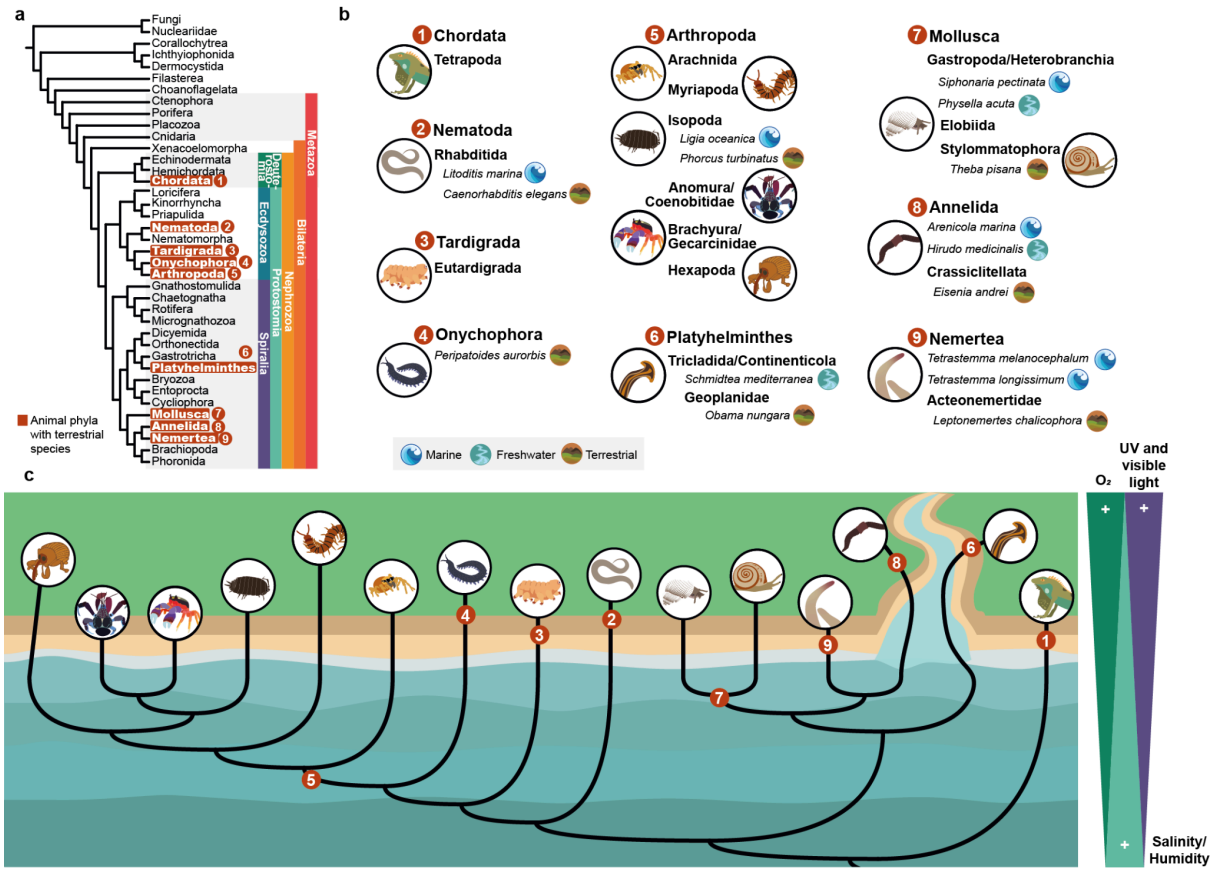


Figure 1. Animal terrestrialisation at a glance. a, Animal tree of life highlighting the phyla where there has been at least one terrestrialisation event. **b**, Main animal terrestrialisation events within animal phyla discussed in this work (in bold), with the species used in the stress

experiments below. **c**, Simplified view of paths towards terrestrial life in animals, and the differences in some abiotic factors between habitats. *[High resolution figure can be accessed here: Figure1.pdf].*

A kingdom-wide reconstruction of animal gene repertoire evolution

To understand the evolutionary foundations of animal terrestrialisation, we first reconstructed the origin and diversification of gene families across the animal kingdom. We performed orthology inference on high-quality protein-coding gene sets from 973 species (960 animals, Table S1), spanning all 31 recognised animal phyla, resulting in over 500,000 orthogroups. Then, we dated the evolutionary origin of each orthogroup by identifying the most recent common ancestor of all the species in which that orthogroup is present and quantified the general patterns of gene gain (first appearance of a given orthogroup, as distinct from gene duplications) and subsequent loss across the metazoan tree (Fig. 2a, S1). Our results showed that most orthogroups originated early in animal evolution, with many orthogroups established during the initial diversification of Metazoa, as previously proposed[30–32]. Notably, and contrary to previous research, we identified a striking burst of gene gain at the base of Protostomia (the lineage that includes all invertebrates that later transitioned to terrestrial environments), probably unveiled by our higher taxonomic sampling (Fig. 2b). In contrast, gene losses were much lower at deeper nodes, followed by independent bursts of losses, which are gradually reduced towards the tips of the tree. These patterns are also reflected in the gene gain-to-loss ratio (Suppl. Fig. S2), where there is a shift from higher gene gain to higher gene loss after Protostomia. While this bias may partly reflect methodological limitations, as gene gain/loss inference under parsimony is sensitive to tree topology, known major phylogenetic discordances in our tree generally predate terrestrialisation events and are unlikely to affect the overall pattern around these transitions. All in all, our results are congruent with previous findings[30–32], and consistent with a biphasic model of evolution[33].

Our extensive dataset allowed us to study 15 independent animal terrestrialisation events represented by two or more species (Fig. 1b), encompassing both ancient transitions deep in the metazoan tree and more recent, lineage-specific events closer to the tips of the phylogeny (i.e., Tetrapoda, Geoplanidae, Acteonemertidae, Crassiclitellata, Ellobiida, Stylommatophora, Rhabditida, Eutardigrada, Onychophora, Arachnida, Myriapoda, Isopoda, Coenobitidae,

Gecarcinidae, and Hexapoda). Compared to older nodes, fewer orthogroups originated at the nodes corresponding to terrestrial lineages. Instead, these nodes were marked by extensive gene loss, in some cases exceeding gene gains by more than an order of magnitude, with gene-to-loss ratios lower than 1 (Fig. 2c, Suppl. Fig. S2, Table S2). This pattern was held consistently across independent terrestrialisation events in and within different phyla. To gain insight into the function of those genes, we then performed separate GO enrichment analysis on orthogroups inferred to be gained or lost at those nodes, using the full genomic gene set of representative species as background (Suppl. Figs. S3-92, see Methods). Genes belonging to both gained and lost orthogroups were significantly enriched in functional terms associated with response to stimuli (such as UV radiation or oxygen), DNA repair and maintenance, reproduction and meiosis, cell division and cell cycle regulation, reproduction, immunity and defense mechanisms, sensory perception and signaling, transport and homeostasis, and developmental processes related to organ formation (e.g., fin development GO term was enriched in lost orthogroups in tetrapods, Suppl. Fig. S90) and environmental adaptation. These functions are exemplified in the gene names that could be obtained using BLASTp. For example, some genes inferred to be gained in tetrapods include the human genes *HEY1*, *AMOT*, *LDHA, COL1A1, TNNT2,* and *TNNT3,* which are related to muscular and cardiovascular function and development[34–39], and might have contributed to the locomotion changes during their water-to-land transition. Moreover, GO terms showed moderate functional convergence (measured as semantic similarity between the GO terms; see Methods) between independent terrestrial lineages (Suppl. Fig. S93a,b). For instance, when we compared hexapods and tetrapods, we found GO terms enriched in functions related to skin protection: cuticle development for hexapods (Suppl. Fig. S3) and keratinization for tetrapods (Suppl. Fig. S87). In addition, some OGs independently gained or lost in at least two terrestrial lineages contain genes with BLASTp hits to the same superfamilies, some of which having been previously linked to terrestrialisation[18–22,40], including CYP450s, mucins, solute carrier proteins, aquaporins, vitellogenin, odorant receptors, or cuticle proteins, among others with more general cellular functions. This suggests a mechanism of paired gain and loss of genes with overlapping functions, reshaping the gene repertoire during terrestrialisation. Overall, the predominance of gene loss over gain at key phylogenetic nodes indicates that, during the transition to land, gene repertoire evolution may have relied more on the removal of functionally redundant or environmentally dispensable genes, facilitated by mutational robustness, rather than relying primarily on de novo gene innovation.

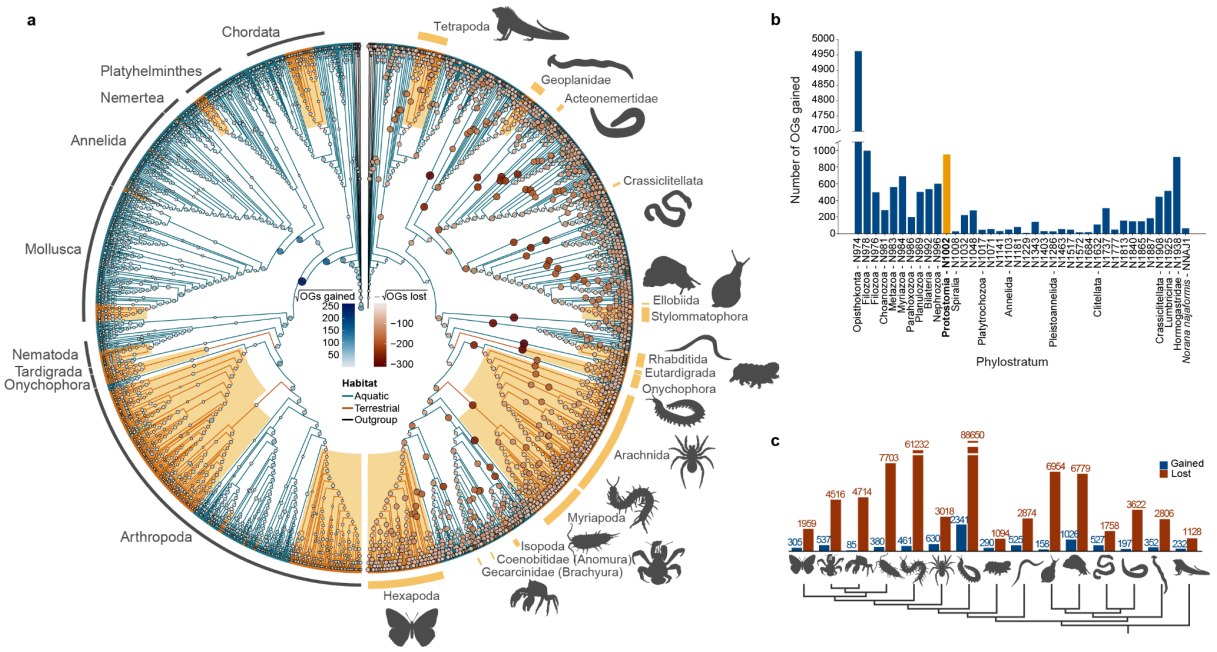


Figure 2. Orthogroup evolution across the animal kingdom. a, Mirror phylogeny of 960 animal species plus outgroups. Circles at each internal branch represent the square-root-scaled number of orthogroups gained (left, blue) and lost (right, brown). Terrestrial lineages are highlighted in yellow; the right phylogeny labels individual terrestrial lineages, while the left phylogeny indicates the phyla containing those lineages. Exact numbers of OGs gained and lost can be found in Suppl. Fig. S1. **b**, Phylostratigraphic origin of all genes in all nodes in a representative species (the earthworm *Norana najaformis,* Annelida), highlighting the gene gain burst in Protostomia (in yellow). **c**, Comparison of gene gain and loss across the 15 terrestrial clades with independent terrestrialisation events. *[High resolution figure can be accessed here: Figure2.pdf].*

Lack of a universal stress-response genomic toolkit for animal terrestrialisation

The transition from aquatic to terrestrial environments exposed animals to a consistent set of abiotic stressors, including water loss and osmotic imbalance, fluctuating oxygen availability, increased ultraviolet (UV) radiation and novel light regimes, as well as requiring changes in chemical signal detection, all of which are widely recognised as major selective pressures during terrestrialisation[24–29]. These stressors are expected to act across phylogenetically diverse lineages, making them suitable candidates for identifying shared or convergent genomic

responses arising from the repeated and independent transition to terrestrial environments by animals.

To test whether a common stress-response genomic toolkit underlies terrestrial adaptation, we designed an experimental framework that directly targets these core abiotic stressors (Fig. 3a). Specifically, we exposed 17 species spanning 7 animal phyla (Nematoda, Onychophora, Arthropoda, Platyhelminthes, Mollusca, Annelida, Nemerta), to controlled stress conditions representing: desiccation and osmotic stress (water loss in terrestrial taxa; salinity shifts in aquatic taxa), oxygen variation (hypoxia and hyperoxia), UV-B radiation with recovery treatments, visible light exposure and reception of chemical cues (Fig. 1b, 3a, Tables S1, S3-S19; see Material and Methods). Following exposure, we generated 796 RNA-seq datasets across all treatments. For each species, RNA material from all experimental conditions was pooled and sequenced using IsoSeq to construct a de novo transcriptome assembly (Table S20), which was used as a reference for mapping RNA-seq reads and gene expression quantification (Tables S3-S19). In parallel, we sequenced control samples using proteomics to validate the translation of expressed proteins.

We then constructed species-specific weighted gene co-expression networks (WGCNA) based on a scale free topology (SFT) (Suppl. Fig. S94-S110, Table S21), and identified statistically significant modules per condition (Suppl. Fig. S111-S127, Tables S22-S38). All genes belonging to these modules were retained for downstream analyses (see Data availability for gene lists) and mapped to the previously inferred orthogroups to identify their evolutionary origin (see Materials and Methods), hereafter referred to as module-associated orthogroups. In addition, we identified highly connected genes (hub genes) within each module, as these are considered to play a central role in response to stress[29–31] (Tables S39-S55). Nonetheless, hubs are defined by network connectivity and are often enriched for essential regulatory genes and the results may be biased toward evolutionary older components[44–47], thus we used them as a complementary dataset to further assess the observed patterns. This framework allowed us to identify orthogroups exhibiting stress-responsive expression patterns across taxa, irrespective of their age of origin, and to test for the presence or absence of a shared genomic toolkit associated with terrestrial adaptation.

The evolutionary origin of module-associated orthogroups was determined using custom scripts based on the idea of phylostratigraphic analysis[48] (Fig. 3b; see interactive link here:

https://itol.embl.de/tree/3722311810722599177072288 5), enabling comparison of gene age distributions across species. To avoid redundancy, each orthogroup was counted only once per species, regardless of how many genes from that orthogroup were present within the significant modules**.** For unbiased cross-species comparison, we calculated the proportion of module-associated orthogroups assigned to each evolutionary branch for each species, based on the total number of module-associated orthogroups identified in that species. We repeated the analysis for genes recovered through the proteomic dataset to assess their translation into proteins. We then compared the proportion of module-associated orthogroups across all experiments and species (Fig. 3c, Table S56) to reveal patterns in stress-response mechanisms associated with the transition to land.

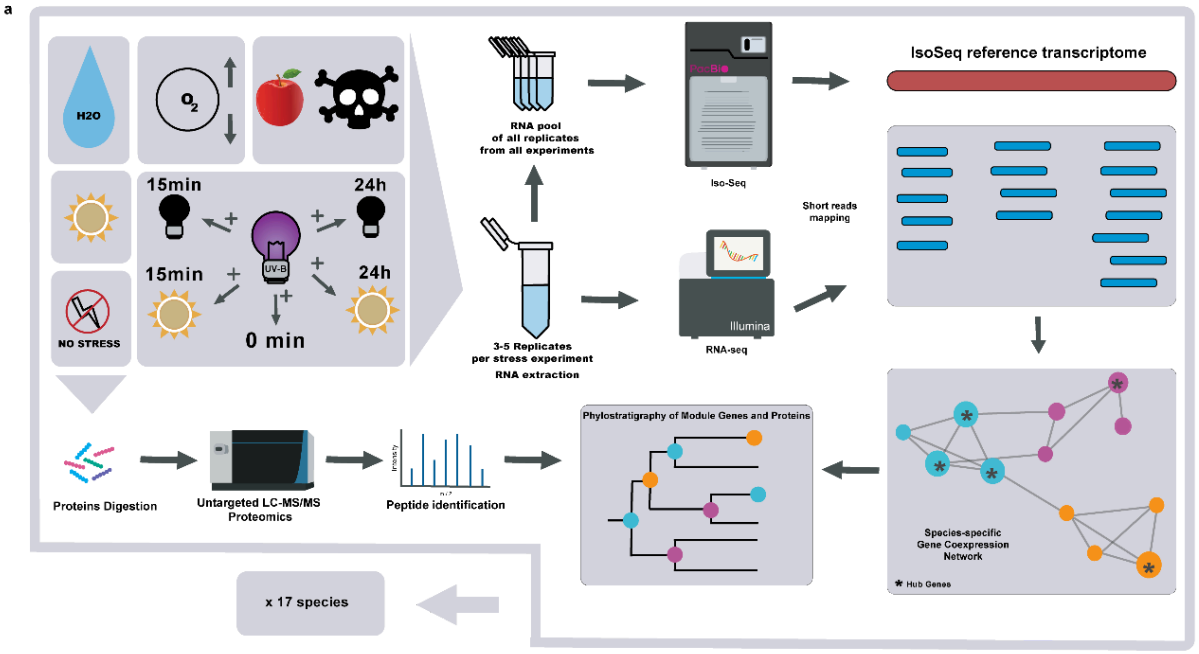
Across all species, the majority of module-associated orthogroups traced back to deep evolutionary nodes, predating the emergence of terrestrial lineages and extending to early metazoan and pre-metazoan origins (Fig. 3b). Only a small fraction of module-associated orthogroups appeared to have arisen more recently, after the origin of Protostomia. While most species followed this general pattern, a few lineage-specific deviations were observed. For example, Onychophora showed an enrichment of module genes in orthogroups gained at the branch leading to the phylum as well as species specific gains; Nematoda exhibited a moderate proportion of module-genes derived from orthogroups gained at the most recent common ancestor of Nematoidea and Panarthropoda (i.e., nematodes, nematomorphs, arthropods, tardigrades and onychophorans); Platyhelminthes showed a cluster of module-genes arising in Tricladida, consistent with prior findings[19]; and the terrestrial nemertean species, along with the terrestrial arthropod species displayed species-specific gains (see interactive link here: https://itol.embl.de/tree/3722311810722599177072288 5. However, even in these cases, the majority of module-genes traced back to orthogroups that predated terrestrialisation and originated in older nodes. Notably, the same pattern was consistent when we limited the analysis in the hub gene set (Suppl. Fig. S128a).

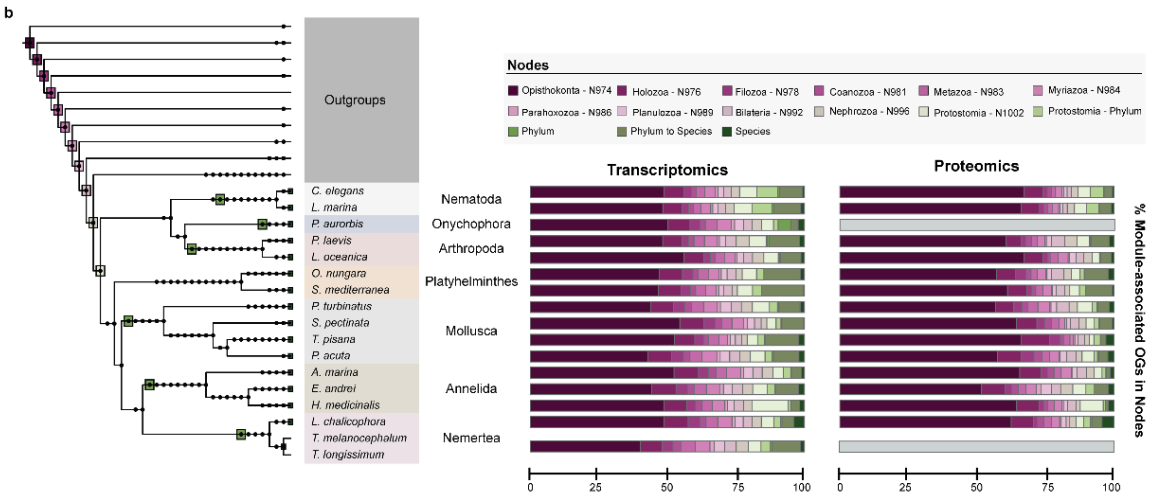
To assess whether different abiotic challenges triggered comparable module-level responses across taxa, we examined the proportion of module-associated orthogroups recovered in each species for each experimental condition (Fig 3c, Table S56). Although the proportion of module-associated orthogroups varied across treatments within the same species, no condition showed a consistent pattern across taxa. All stressors revealed strong responses in some

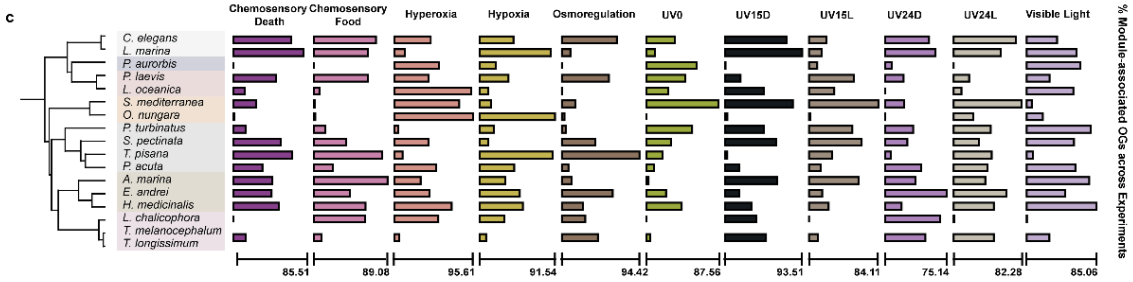
species, but weak in others. This indicates that there is not a common stress-specific signature, rather than an heterogeneous, lineage-dependent employment of stress-responsive modules.

While the evolutionary origins of stress-associated genes were broadly conserved, the specific orthogroups recruited into stress-response networks were highly variable. Only 42 module-associated orthogroups out of the ca. 540,000 were shared across all 17 species (Fig. 3d), encoding proteins conserved across eukaryotes and with previously identified stress-related roles, such as RAB proteins[32], selenoproteins[33], ubiquitin carboxyl-terminal hydrolase 7[34], tetraspanin[35,36], cullin[37], or putative voltage-dependent anion-selective channel[38] (Table S57). When considering the hub genes of the network, we observed virtually no overlap in the orthogroups to which hub genes belonged - not among terrestrial species, not among aquatic species, and not even within the same phylum, except for nematodes (Suppl. Fig. S128b). These findings indicate that stress responses have largely evolved through independent recruitment of distinct gene repertoires across lineages, rather than relying on a conserved genomic toolkit, though a small set of deeply conserved genes may underpin core functions across animals.

These results argue against a universal genomic toolkit for response to abiotic factors relative to life on land. The lack of orthogroup-level overlap among the module-associated genes recruited into stress-response networks suggests that different lineages independently co-opted distinct sets of genes to meet the environmental pressures associated with transition to land. Importantly, this pattern holds across both terrestrial and aquatic species and both within and between phyla, underscoring the absence of a conserved genetic solution to stress. Instead, response to abiotic stress appears to have proceeded through the lineage-specific reorganization of deeply conserved orthogroups into different genetic contexts. Overall, our findings support the idea that gene reuse diminishes with increasing evolutionary divergence[49], even in response to similar abiotic stressors.


a
H2O
O2
15min
24h
15min
24h
UV-B
0 min
NO STRESS
RNA pool
of all replicates
from all experiments
Iso-Seq
IsoSeq reference transcriptome
Short reads
mapping
3-5 Replicates
per stress experiment
RNA extraction
RNA-seq
Illumina
Proteins Digestion
Untargeted LC-MS/MS
Proteomics
Peptide identification
Phylostratigraphy of Module Genes and Proteins
Species-specific
Gene Coexpression
Network
Hub Genes
x 17 species


b
Outgroups
C. elegans
L. marina
P. aurorbis
P. laevis
L. oceanica
O. nungara
S. mediterranea
P. turbinatus
S. pectinata
T. pisana
P. acuta
A. marina
E. andrei
H. medicinalis
L. chalicophora
T. melanocephalum
T. longissimum
Nodes
Opisthokonta - N974
Holozoa - N976
Filozoa - N978
Coanozoa - N981
Metazoa - N983
Myriazoa - N984
Parahoxozoa - N986
Planulozoa - N989
Bilateria - N992
Nephrozoa - N996
Protostomia - N1002
Protostomia - Phylum
Phylum
Phylum to Species
Species
Transcriptomics
Proteomics
Nematoda
Onychophora
Arthropoda
Platyhelminthes
Mollusca
Annelida
Nemertea
0
25
50
75
100
% Module-associated OGs in Nodes


c
Chemosensory Death
Chemosensory Food
Hyperoxia
Hypoxia
Osmoregulation
UV0
UV15D
UV15L
UV24D
UV24L
Visible Light
C. elegans
L. marina
P. aurorbis
P. laevis
L. oceanica
S. mediterranea
O. nungara
P. turbinatus
S. pectinata
T. pisana
P. acuta
A. marina
E. andrei
H. medicinalis
L. chalicophora
T. melanocephalum
T. longissimum
85.51
89.08
95.61
91.54
94.42
87.56
93.51
84.11
75.14
82.28
85.06
% Module-associated OGs across Experiments

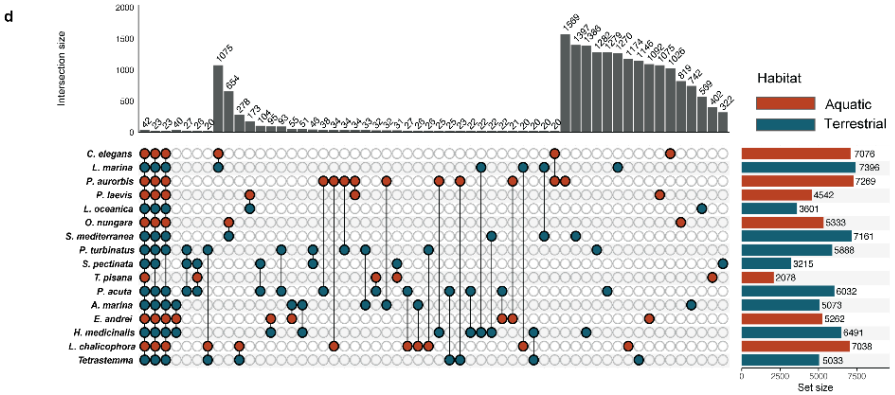

d
Intersection size
2000
1500
1000
500
0
1075
604
278
73
1566
1397
1366
1282
1276
1270
1174
1146
1092
1075
1066
919
842
589
492
322
Habitat
Aquatic
Terrestrial
C. elegans
L. marina
P. aurorbis
P. laevis
L. oceanica
O. nungara
S. mediterranea
P. turbinatus
S. pectinata
T. pisana
P. acuta
A. marina
E. andrei
H. medicinalis
L. chalicophora
Tetrastemma
7076
7396
7269
4542
3601
5333
7161
5888
3215
2078
6032
5073
5262
6491
7038
5033
0
2500
5000
7500
Set size

Figure 3. Species-specific stress-response networks and pre-metazoan evolutionary origins of stress-related genes suggest diverse genomic routes to terrestrial life in animals. **a**, Overview of abiotic stress experiments, co-expression network construction and proteomic detection of stress-related genes across 17 species from 7 animal phyla, exposed to standardised stressors mimicking terrestrial abiotic factors. **b**, Phylostratigraphic analysis showing that the majority of stress-related genes belong to old orthogroups, based on genes recovered through both transcriptomic and proteomic data. Colored squares in the tree indicate the branch leading to each node, using the same color scheme as in the accompanying barplot. Nodes from collapsed branches not shown in the tree are marked with black dots. Protostomia to Phylum and Phylum to Species segments indicate the sum of all the in-between branches. Grey bars indicate missing data for specific species. **c**, Percentage of module-associated orthogroups per species for each experiment, calculated relative to the total number of module-associated orthogroups identified in that species. **d**, Upset plot showing overlap of module-associated orthogroups resulted from species-specific co-expression networks. Dots connect species that have a number of module-associated orthogroups in common, the number indicated in each bar from the barplot on top. *[High resolution figure can be accessed here: Figure3.pdf].*

Functional convergence in species-specific module genes in response to abiotic stress

In the absence of shared orthogroups among stress-response co-expression networks across species, we next asked whether convergence occurs at the level of gene function. To explore this, we applied pathway comparisons of statistically significant modules of each species and we investigated the functional similarity of hub genes identified in co-expression networks from each species exposed to abiotic stress, applying gene ontology-based comparisons, following the approach described in[18,50] (see Materials and Methods).

First, we examined pathway-level convergence by assessing enrichment of well-defined functional categories. Despite species-specific gene repertoires, the pathways enriched in module gene sets under each stress condition were largely shared across species, indicating parallel recruitment of distinct gene sets into common functional categories. When considering all stress conditions together, a core set of conserved pathways emerged as recurrently enriched across lineages (Fig. 4a, Suppl. Fig. S129a). Pathways present in more than 80% of species

included central metabolic processes (e.g., carbon and purine metabolism), genetic information processing pathways (e.g., spliceosome, ribosome, nucleocytoplasmic transport, and protein processing in the endoplasmic reticulum), as well as major signaling cascades including MAPK, PI3K–Akt, AMPK, and mTOR signaling. In addition, pathways associated with cellular stress responses (such as endocytosis, lysosome, autophagy, apoptosis, and cellular senescence) together with processes maintaining cellular architecture (focal adhesion, tight junctions, and cytoskeletal regulation) were consistently recovered across species (Fig. 4a). This recurrence of the same functional categories across phylogenetically distant species and diverse environmental conditions indicates that metazoan responses to environmental stress rely on a deeply conserved molecular toolkit. Such convergence is consistent with the concept of a universal cellular stress response, in which evolutionarily ancient pathways are repeatedly co-opted to maintain cellular homeostasis[51–53]. Nonetheless, the specific pathways activated varied depending on both the species and the type of stressor (right-side heatmaps in Fig. 4a, Suppl. Fig. S129a), suggesting that different lineages mobilize distinct components of these conserved functional modules in response to specific environmental challenges. This variation did not follow a consistent phylogenetic or ecological pattern, and is therefore likely to reflect differences in transcriptional responses (see Suppl. Figs. S130-S140 for unfiltered pathway lists per experimental condition). However, further resolution of these patterns revealed that convergence is not limited to pathway presence, but also extends to pathway prominence. The most frequently represented pathways within each experiment showed that a consistent set of pathways, including metabolic pathways, PI3K–Akt and MAPK signaling, endocytosis, focal adhesion, and components of gene expression and proteostasis, were repeatedly among the most dominant across conditions (Fig. 4b, Suppl. Fig. S129b). This indicates that stress responses are organized around a small number of core regulatory modules that are consistently prioritized across species and environmental contexts.

To further assess the functional organization of stress-responsive modules, we examined their pathway composition across species. We found that individual modules were typically associated with multiple pathways rather than being dominated by a single functional category. (Table S58). This indicates that co-expression modules are not pathway-specific units but instead represent integrated functional assemblies encompassing multiple biological processes[54].

To complement these results, we investigated whether core-regulatory elements (i.e., hub genes) are also showing consistent signatures of functional convergence, this time at the gene ontology (GO) level. In brief, we inferred GO annotation similarity across all hub-containing orthogroups per species, measured their functional similarity and grouped them into functional clusters that eventually represent functional convergence units, which we represent in what we name as constellation plots[18,50] (Fig. 5a). For each experiment, we analyzed the full hub gene set of each species in an aggregated manner and we found that hub genes across different lineages shared highly similar functional annotations, though they belonged to entirely different orthogroups. When grouped by functional similarity, hub-containing orthogroups from different species clustered into one to four functional clusters including both aquatic and terrestrial animals, with most species always being included in a single main cluster (Fig. 5b, Fig. S141). These results suggest convergent recruitment of genes with analogous roles in abiotic stress response despite divergent evolutionary origins. This pattern was further supported by orthogroup-level clustering, which revealed that the vast majority of functional clusters comprised orthogroups containing both aquatic and terrestrial species, with only a minority of clusters showing habitat-specific composition (Tables S59-S69).

Together, these results reveal that, while the genetic content of stress-response networks differs between species, their functional roles converge across lineages. This indicates that animals transitioning to land did not reuse the same genes, but rather recruited distinct orthogroups (most of them with a pre-metazoan evolutionary origin) to produce similar functional outcomes. Such functional convergence, emerging from non-homologous gene sets, underscores the evolutionary flexibility of gene networks and suggests that response to stress during terrestrialisation was achieved through multiple genomic trajectories leading to comparable physiological outcomes.

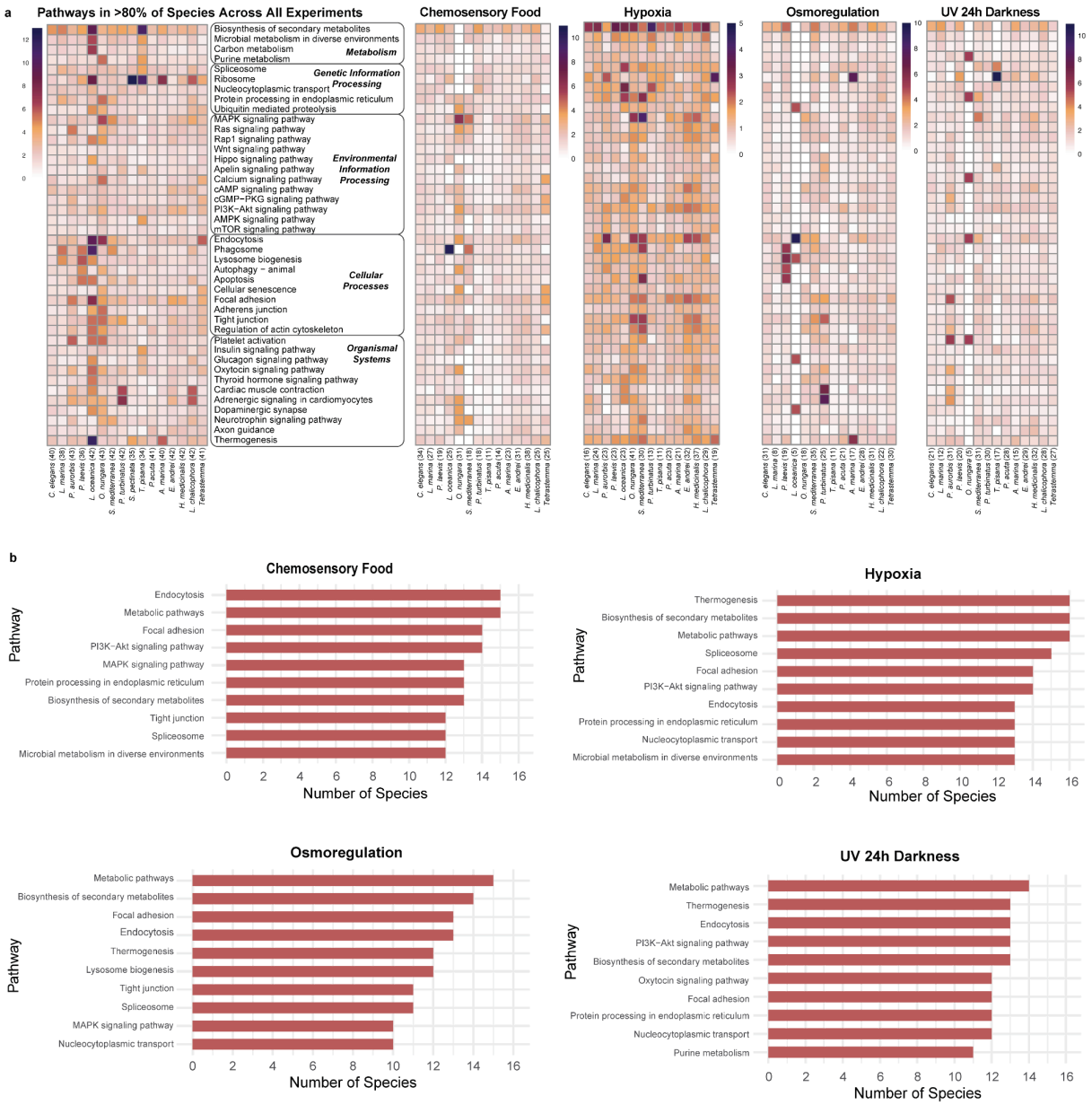


Figure 4. Functional convergence of species-specific KEGG pathways across animal lineages under stress conditions. a, Heatmaps showing the proportion of KEGG pathways in gene sets under abiotic stress conditions**.** Each row represents a KEGG pathway, and each column corresponds to a species. The left heatmap displays a summary heatmap including only pathways that are present in at least 80% of species, considering all experimental conditions combined. On the right, separate heatmaps are shown for each experimental condition, filtered by the same set of pathways (rows) as in the summary heatmap. Only four experiments are shown; all experiments showed the same pattern (see also Suppl. Fig. S129a). Numbers in parentheses next to species names indicate the total number of enriched pathways detected in that species for each condition. Color scales represent the percentage of genes in each pathway.

b, Top 10 pathways per experiment. The same four experiments as in 4a are shown (see also Fig. S129b). *[High resolution figure can be accessed here: Figure4.pdf].*

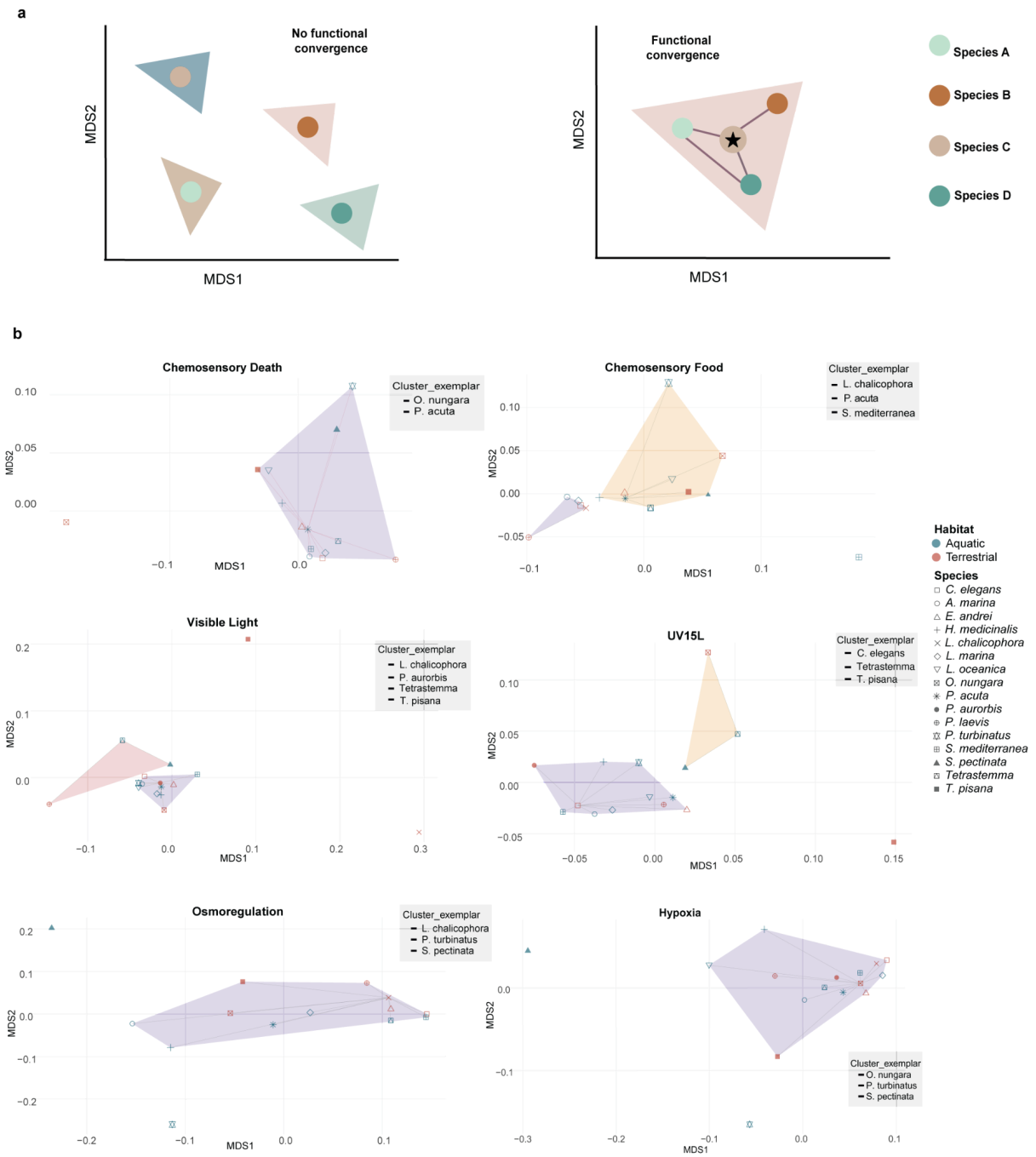


Figure 5. Functional convergence of species-specific, GO-annotated hub genes across animal lineages under stress conditions. a, Graphical illustration of constellation plots of gene functions across species. Each dot represents a species and each polygon a distinct functional cluster. On the left side each species belongs to a different cluster showing the lack of functional convergence and on the right side all species cluster together representing

functional convergence. **b**, Constellation plots of functions of hub genes across 17 aquatic and terrestrial species for each experimental condition. Each polygon represents a distinct functional cluster formed by species with similar GO annotations, despite no orthogroup-level overlap across species. Cluster exemplars indicate the species that act as centroid in each cluster. The presence of hub genes from multiple species within the same functional cluster demonstrates convergence in stress-response roles across independent lineages. Only six experiments are shown for demonstration purposes, given that all experiments showed the same pattern (see also Suppl. Fig. S141). Functional convergence measured as GO-annotation similarity was inferred by calculating Wang semantic similarity between species' GO term sets using the Best-Match Average method, followed by clustering with Affinity Propagation. *[High resolution figure can be accessed here: Figure5.pdf].*

Machine learning identifies habitat-informative gene families across animal phyla

Although our stress-response experiments provide insight into how animals respond to abiotic challenges, they capture only part of the biological complexity involved in terrestrialisation, excluding major factors such as anatomical, embryological, biomechanical, or biochemical adaptations. To explore broader genomic signals potentially linked to the terrestrial habitat, we examined whether gene copy number variation could predict habitat. Gene copy number variation is a key mechanism of evolutionary adaptation, often reflecting functional shifts linked to ecological transitions[18,50,55] that often underpin physiological and functional innovations. To test this, we applied machine learning, a method that has successfully identified environment-related genes in previous studies[56,57]. We trained an XGBoost machine learning classifier using our dataset of over 500,000 orthogroups inferred for our nearly 1,000 proteomes spanning all major animal phyla, providing a kingdom-wide assessment of putative habitat-associated orthogroups (Fig. 6a).

Our model identified a small subset of orthogroups (130 orthogroups, Table S70) as the minimum number of orthogroups that accurately distinguish terrestrial from aquatic species (Fig. 6b,c). Orthogroup count distributions along the phylogeny showed that our model classified species based on complex patterns of copy number variation not easily assessed by eye (Suppl. Fig. S142-271). For example, orthogroup (OG) 8389 (Selenoprotein L, see below) is almost exclusively found in aquatic organisms (Suppl. Fig. S240), while OG492

(lysosomal acid phosphatase) shows a higher count in some terrestrial lineages than their aquatic relatives, but the opposite trend in others (Suppl. Fig. S171). This suggests that complex gene duplication and loss dynamics of genes originating before terrestrialisation events could have contributed to the transition of animals to life on land. Importantly, the discriminative power of these orthogroups does not arise from simple presence or absence or congruent expansions and contractions in independent terrestrial lineages, but from lineage-specific shifts in orthogroup size, reflecting diverse, independent, and complex evolutionary trajectories that led to convergent adaptation to terrestrial environments.

Functional annotation of these habitat-informative orthogroups revealed gene families involved in processes relevant to environmental stress, metabolism, signaling, and development (Fig. 6d, Table S70), comprising metabolic enzymes, ion and solute transporters, and odorant receptors, which share functional roles with, or are themselves, gene families previously linked to terrestrialisation[18–22,40], among others. Known gene families include the cytochrome P450s (CYP450s) and flavin-containing monoxygenases (FMOs), involved in the detoxification of xenobiotics, cellular metabolism, and homeostasis. Both were parallelly expanded in independent mollusc out-of-the-sea transitions[18]. CYP450s were also previously associated with critical functions during plant transition to land[58], while FMOs are involved in the metabolism of Trimethylamine[59], characterised by its strong fish-like odor. OG6 (serine proteases) includes vertebrate coagulation factors, one of which (factor XII) emerged before the appearance of lungfishes, during the tetrapod water-to-land transition, and has been hypothesised to confer adaptive advantages by enabling rapid clotting in soil-contaminated wounds, mitigating infection and blood loss in the terrestrial environment[60]. Other less-studied gene families identified as relevant for animal terrestrialisation were Rapunzel-like, which participates in skeletal and muscular development in fish but is absent in tetrapods[61], and Selenoprotein L, whose function remains unknown, but was predicted in our functional analysis to be involved in chromosome segregation. The latter protein incorporates the rare amino acid selenocysteine, which depends on environmental selenium for its synthesis. It has been hypothesised that the greater bioavailability of selenium in aquatic environments, along with the increased oxidative stress associated with higher oxygen levels on land, has led aquatic species to retain larger selenoproteomes than their terrestrial counterparts[62,63], highlighting a potential link between terrestrialisation, oxygen availability, and selenoprotein evolution. Transport of ions and solutes, such as trehalose, involved in osmoregulation, was also represented in the identified gene families. These functions are broadly consistent with the

physiological demands of land life, including coping with oxidative damage, managing dehydration and oxygen variability, and processing novel chemical cues.

We further evaluated the overlap between those habitat-informative orthogroups and the hub genes previously identified in species-specific co-expression networks derived from transcriptomic experiments. Out of the 130 total orthogroups identified through machine learning, 80 contained hub genes from at least one species (Fig. 6a). This overlap suggests that gene families whose copy number variation correlates with habitat may also contribute to real-time transcriptional responses to environmental challenges. However, these orthogroups account for only a subset of the total hub gene repertoire, and their involvement as hubs is not consistent across species, even when the gene families are broadly conserved, underscoring the role of species-specific genomic rewiring in shaping stress responses.

Taken together, these results highlight a small but functionally relevant set of gene families that have undergone recurrent copy number shifts associated with habitat transitions, reflecting parallel evolutionary responses to ecological change. Moreover, their dual roles as predictive features of habitat and as transcriptional hubs suggest that these genes represent flexible genomic elements that might have been co-opted into the physiological response to stress or other challenges posed by terrestrial life.

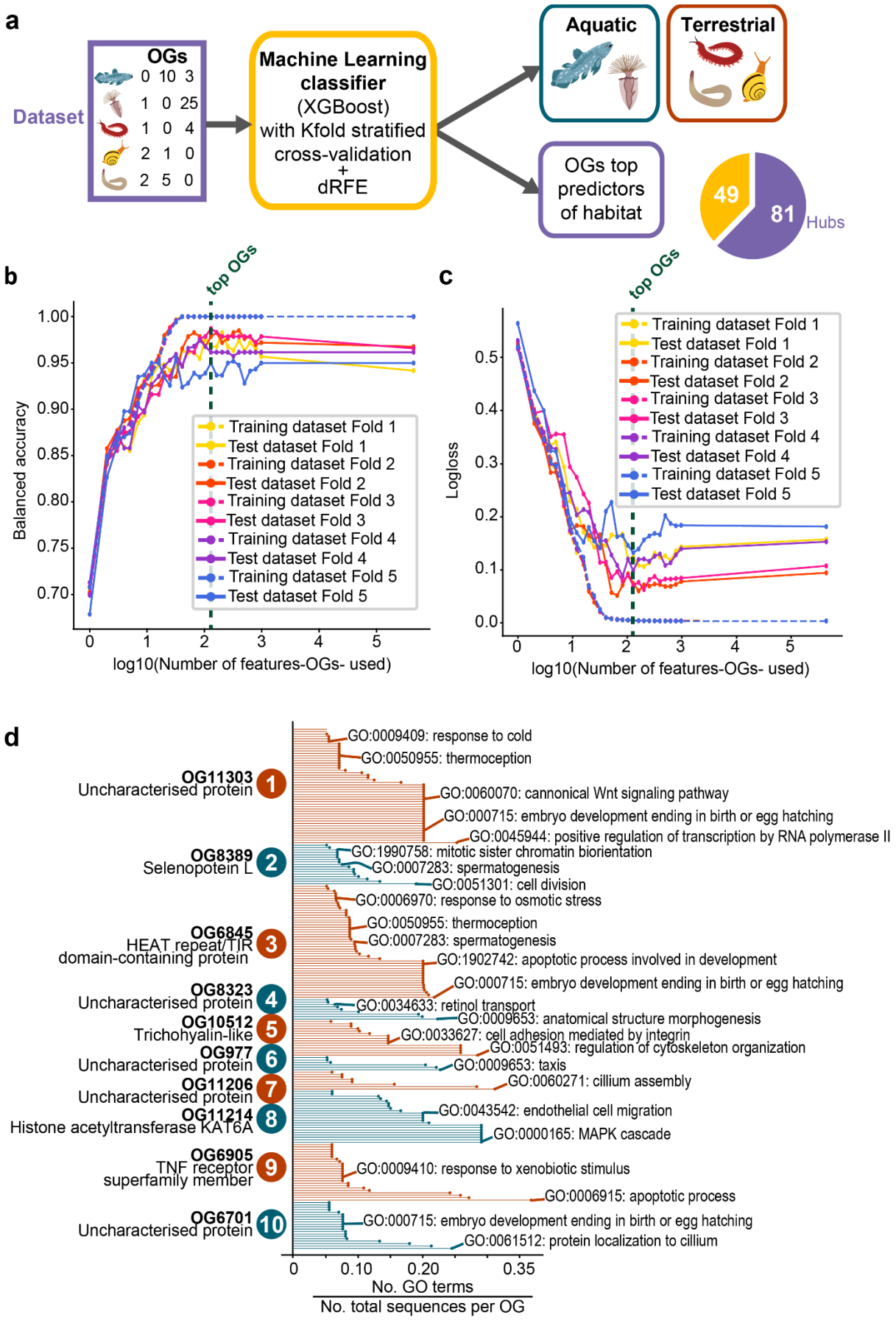

a
Dataset
OGs
0 10 3
1 0 25
1 0 4
2 1 0
2 5 0
Machine Learning classifier
(XGBoost)
with Kfold stratified cross-validation
+
dRFE
Aquatic
Terrestrial
OGs top predictors of habitat
49
81
Hubs
b
top OGs
Balanced accuracy
log10(Number of features-OGs- used)
Training dataset Fold 1
Test dataset Fold 1
Training dataset Fold 2
Test dataset Fold 2
Training dataset Fold 3
Test dataset Fold 3
Training dataset Fold 4
Test dataset Fold 4
Training dataset Fold 5
Test dataset Fold 5
c
top OGs
Logloss
log10(Number of features-OGs- used)
d
OG11303 Uncharacterised protein
OG8389 Selenopotein L
OG6845 HEAT repeat/TIR domain-containing protein
OG8323 Uncharacterised protein
OG10512 Trichohyalin-like
OG977 Uncharacterised protein
OG11206 Uncharacterised protein
OG11214 Histone acetyltransferase KAT6A
OG6905 TNF receptor superfamily member
OG6701 Uncharacterised protein
GO:0009409: response to cold
GO:0050955: thermoception
GO:0060070: cannonical Wnt signaling pathway
GO:000715: embryo development ending in birth or egg hatching
GO:0045944: positive regulation of transcription by RNA polymerase II
GO:1990758: mitotic sister chromatin biorientation
GO:0007283: spermatogenesis
GO:0051301: cell division
GO:0006970: response to osmotic stress
GO:0050955: thermoception
GO:0007283: spermatogenesis
GO:1902742: apoptotic process involved in development
GO:000715: embryo development ending in birth or egg hatching
GO:0034633: retinol transport
GO:0009653: anatomical structure morphogenesis
GO:0033627: cell adhesion mediated by integrin
GO:0051493: regulation of cytoskeleton organization
GO:0009653: taxis
GO:0060271: cillium assembly
GO:0043542: endothelial cell migration
GO:0000165: MAPK cascade
GO:0009410: response to xenobiotic stimulus
GO:0006915: apoptotic process
GO:000715: embryo development ending in birth or egg hatching
GO:0061512: protein localization to cillium
No. GO terms
No. total sequences per OG

Figure 6. Machine learning identifies 130 orthogroups (OGs) predictors of habitat. a, Simplified workflow of machine learning analysis (full workflow in Suppl. Fig. S272d). The pie chart shows the number of OGs from the top 130 that contain hubs from stress-related modules in GCN. **b**, Balanced accuracy for each model trained across all iterations of the dRFE workflow for training and testing datasets of all folds. The dashed dark green line shows the iteration where we select the top OGs. **c**, Logloss values for each model trained across all iterations of the dRFE workflow for training and testing datasets of all folds. The dashed dark green line shows the iteration where we select the top OGs. **d**, Functions from the top 10 OGs as ranked by the machine learning model. OG name was obtained with BLASTp. GO terms on the right were predicted using FANTASIA. Full GO terms in Suppl. Fig. S273, Table S70. *[High resolution figure can be accessed here: Figure6.pdf].*

Discussion

The repeated transition of animal lineages from an aquatic to a terrestrial environment is one of the most impactful examples of convergent evolution. It required the independent acquisition of physiological and morphological adaptations to cope with novel environmental pressures in terrestrial habitats[28,29]. However, whether these independent adaptations might have been grounded on shared or lineage-specific genomic changes has never been addressed. Our results provide evidence that terrestrial adaptation is characterised by both convergent and parallel (different and same gene families, respectively) reshaping of terrestrial lineages' genomic repertoires. This includes (1) independent gains and losses of genes with similar functions; (2) a lack of shared genomic responses to the same key abiotic stressors encountered during terrestrialisation, concordant with the expectation of reduced gene reuse as divergence times increase[49], and coupled with evidence for pathway and functional convergence[64] ; and (3) copy number variation in conserved orthogroups gained pre-terrestrialisation, primarily involved in stress responses, metabolism, and development, that might have contributed to terrestrial adaptation through complex phylogenetic patterns of duplication and loss.

Our results suggest that the evolution of terrestrial lineages relies mostly on the inherent plasticity and modularity of the pre-existing gene repertoire and molecular networks[64,68]. This contrasts with cases where adaptation relies on novel genes, such as stress response in yeasts[65], or the repeated modification of the same genetic elements, such as the parallel loss of an enhances causing pelvic reduction in stickleback fish[66]. Nonetheless, it is in line with the

classical idea of "evolution as a tinkerer"[67], which can be shaped by[68] or promote the adaptation[71–73] to the ecological pressures of terrestrial life. Previous studies have shown this trend on plant[41–43,70], vertebrate[21], mollusc[18], and flatworm[19] terrestrialisation, and on the response to desiccation in gecarcinid land crabs[69]. This, together with our results, highlights the importance of considering systems-level properties, such as co-expression network architecture, hub gene centrality, and pathway redundancy, when investigating the genomic basis of adaptation and convergent evolution[68]. Importantly, while this plasticity likely reflects intrinsic properties of ancestral gene repertoires and networks, its evolutionary outcome is expected to be context-dependent, with distinct biotic and abiotic conditions across independent terrestrialisation events shaping which components are co-opted, rewired, or lost.

Despite the breadth of our taxonomic and experimental framework, several limitations inherent to the available data and to deep-time comparative inference should be considered when interpreting our results and their generality. Firstly, although our wide taxonomic representation reduces the chance of missing information by incorporating relevant but often overlooked lineages (e.g., nemerteans), it is still biased towards lineages with a high representation in the genomic databases, limiting our confidence on the gained and lost OGs in some terrestrial lineages (e.g. our dataset only contains one terrestrial amphipod, which led us to not consider this terrestrialisation event for further analyses as we cannot distinguish between species and lineage-specific OGs). Secondly, it is not known how many terrestrialisation events occurred in animals and its inference changes depending on the species phylogeny. Perhaps the most clear example is the arachnid terrestrialisation, where the inclusion of Xiphosura within arachnids may challenge the standing idea of a single terrestrialisation event[74], but secondary transitions to aquatic environments followed by transitions back to terrestrial habitats (e.g. in turtles[75]) can complicate the outlook even more. Further palaeontological research on invertebrates may identify additional terrestrialisation events and their ecological context, providing additional information to understand these habitat transitions. Thirdly, the separation between aquatic and terrestrial is not binary. Even though we considered terrestriality as a binary trait to make it comparable between different lineages, it could be argued that some lineages acquired more profound adaptations that allowed them to thrive in fully terrestrial habitats (e.g. arthropods that live on deserts) compared to other lineages where there is still some dependence on water (e.g. terrestrial planarians live in humid environments). A gradual classification with intermediate habitats similar to the one proposed for land crabs[76] cannot be applied when comparing between

lineages due to their profound morphological and physiological differences. Fourthly, terrestrial environments are highly heterogeneous, and the nature and intensity of abiotic stressors have likely varied substantially over evolutionary time. As such, the experimental conditions used here are best interpreted as controlled proxies for broad environmental axes (e.g. water balance, oxygen availability, UV radiation), rather than direct recreations of ancestral terrestrial settings. Furthermore, although we targeted a set of core, experimentally tractable stressors, other important factors such as temperature variation and mechanical constraints were not included and may also have contributed to terrestrial adaptation. Together, these limitations highlight that our results capture how extant taxa respond to terrestrial-like challenges, and should be interpreted as evidence of recurrent functional responses rather than direct reconstructions of ancestral molecular states.

While our present-day stress experiments and phylogenomic approach cannot recapitulate the full complexity of terrestrialisation, they provide a kingdom-wide comparative framework to understand how distantly related animal lineages achieve similar functional adaptations to terrestrial environments through the independent recruitment of distinct gene repertoires. Moreover, as computational tools and global biodiversity genomic initiatives bridge the gap between non-model and model organisms, experimental research on previously neglected lineages will become easier and will provide us with functional information to corroborate or improve computational predictions of gene function. This will benefit more in-depth studies on independent terrestrial lineages, where physiological and molecular information could be combined to get a more comprehensive view of each terrestrialisation event. All in all, our results offer a first molecular glimpse on how animals could have repeatedly overcome the challenges of novel terrestrial environments not only by inventing entirely new solutions, but mostly by tinkering pre-existing ones.

Materials and Methods

Species selection for kingdom-wide gene repertoire evolutionary dynamics

We selected 960 animal species covering all animal phyla, and 13 outgroup species (Table S1). We paid special attention to their taxonomic distribution to ensure that internodes leading to terrestrial lineages were well represented. Per-gene protein sequences were obtained from MATEdb2[79]. For each species, the longest isoform per gene was parsed for orthology

inference. Reference transcriptomes from species collected to carry out abiotic stress experiments were added as well (see below). Information on the source data type for each species (i.e. genome, RNA-seq transcriptome or IsoSeq transcriptome) can be found in Table S1. We grafted our species tree based on the information available in the literature.

Species tree topology sources

Deep relationships between animal phyla are still the topic of debate in the scientific community. For our species selection, we decided to use the most recent and sound phylogenomic studies whenever possible. The use of alternative topologies to the one considered here for these unresolved relationships would not affect the general conclusions of our results, as terrestrialisation events occurred later and most orthogroups that we discussed originated before these events (regardless of the specific timing of origin).

For the most recent interrelationships, there were some cases where only a few markers or even morphological studies that include those species existed. There were a few cases where information on the exact relationship between three or four species was not available (i.e., Talitridae (Arthropoda), *Siphonaria* (Mollusca), Nerididae (Annelida), Opilioacarida (Arthropoda), Sesarmidae (Arthropoda), Hyppolytidae (Arthropoda)), forming a polytomy that was resolved using ETE v3.1.2[80]. These should not affect the conclusions, either, as our results suggest a higher importance for genes that predate these branches.

For several terrestrial lineages (e.g., Talitridae, Amphipoda), only genomic or transcriptomic data of a single terrestrial species were available. This would affect our results in the sense that we would not be able to distinguish between lineage- and species-specific orthogroups. For this reason, these cases were not considered for the gene repertoire evolution analyses.

Methods Table 1. Species phylogeny topology sources per lineage/group.

Lineage/clade/group of species	Reference(s)
Outgroups	81
Metazoa backbone	82

Lineage/clade/group of species	Reference(s)
Porifera	[83,84,85]
Cnidaria	[86,87]
Acoela	[88]
Echinodermata	[89]
Hemichordata	[90]
Cephalochordata	[91]
Craniata Actinopterygii Amphibia	[92,93] [94–96] [97]
Ecdysozoa	[98,99]
Kinorhyncha	[100]
Nematoda *Litoditis* + *Oscheius* Tylenchina + Panagrolaimidae	[101] [102] [103]
Nematomorpha	[104]
Tardigrada Macrobiotidae Richtersiidae	[105–108] [109] [110]
Onychophora *Eoperipatus*	[111] [112]
Arthropoda Chelicerata Astigmata Trombidiformes	[113] [114] [115] [116]

Lineage/clade/group of species	Reference(s)
Pseudoscorpiones	117
Araneae	118
Myriapoda	119
Lithobiomorpha	120
Scutigeromorpha	121
Pancrustacea	122
Oligostraca + Malacostraca + Branchiopoda	123
Myodopida + Podocopida	124
Podocopida +Cypridocopida	125
Notostraca + Spinicaudata	126
Lepidurus triops	127
Malacostraca	122
Amphipoda	128
Pardaliscidae	129
Tanaidacea	130
Isopoda	131,132
Asellidae	133
Paraplatyarthrus subterraneus	134
Euphausiacea	135
Decapoda	136
Caridea	137
Sergestidae	138
Anomura	139
Brachyura	140
Thoracotremata	141
Hexapoda	142
Collembola	143
Entomobryomorpha	144
Tomocerinae	145
Diplura	146
Protura	147

Lineage/clade/group of species	Reference(s)
Insecta	142
Blattodea	148
Paraneoptera	149
Coleoptera	150
Lepidoptera	151
Hymenoptera	152
Diptera	153
Protostomia	
Cycliophora	154
Dicyemida	155
Spiralia	156
Chaetognatha	157
Rotifera	158
Platyhelminthes	159,160
Bryozoa	161,162
Entoprocta	163
Mollusca	164,165
Polyplacophora	166
Caudofoveata + Solenogastres	167
Cephalopoda	168–170
Gastropoda	171
Neomphalida	172
Vetigastropoda	173
Patellogastropoda + Neritimorpha	174–176
Caenogastropoda	177,178
Nudibranchia	179
Panpulmonata	180

Lineage/clade/group of species	Reference(s)
Helicidae	181
Arion vulgaris + *Meghimatium*	182
Ampullaridae	177
Bivalvia	183
Unionida	184
Mytilida + Ostreoidea	185
Ostreidae	186
Annelida	
Early-splitting annelids + Sedentaria	187
Siboglinidae	188
Oasinia alvinae	189
Spinther	190
Sipuncula	191
Errantia	192
Aciculata	193
Dimorphilus + Sedentaria	194
Sabellida	195
Terebelliformia	196
Clitellata	197
Whitmania pigra	198
Nemertea	199
Hoplonemertea	200
Lineidae	201
Brachiopoda	202,203
Phoronida	204

Extant and ancestral habitat reconstruction

Habitat information for each species was mostly obtained from the World Register of Marine Species (WoRMS) database[205] (last accessed in January 2023). To simplify, a species was considered terrestrial if it spent at least one stage of its life cycle in a terrestrial environment. The only exceptions were the diving bell spider, *Argyroneta aquatica*, and the maritime centipede, *Strigamia maritima,* which were considered terrestrial because they conserve most adaptations to life on land.

We inferred the ancestral habitat of each branch using the R package paleotree v3.4.7[206] using the MPR method, which provided a reconstruction more congruent with the literature than the ACCTRAN method. For ambiguous and wrongly assigned branches (Suppl. Fig. S274), we manually assigned the habitat to aquatic (Suppl. Fig. S275). Given that information on where terrestrialisation events happened is already described in the literature, and that a comprehensive testing of whether they occurred in those moments requires an even wider taxon sampling, that also incorporates fossil information (if available), and more powerful methods that cannot be applied here, the only purpose of this analysis was to map, in an automated manner, the information from the literature in the main figures. Results were not used for any additional analyses.

MPR gives ambiguous reconstructions in 3 lineages: nematodes, molluscs, and annelids. For nematodes, no paper infers the ancestral habitat state. One paper suggested more than 30 transition events (Holterman et al., 2019), but they did not distinguish between terrestrial and freshwater. The ambiguous group in question (Plectida) includes terrestrial and aquatic (marine) clades (supplementary figure 1 in that paper), making it more likely that the ancestor was aquatic. More specifically, we can say that branch 1116 from our tree is most likely aquatic, while 1117 could also be terrestrial. For molluscs, we used the habitats previously inferred[180]. For annelids, we used the habitats previously inferred[197]. One species (*Drawida*) was considered as part of Crassiclitellata in their ancestral reconstruction, but not in their final selected topology. So, we considered this as a putative secondary terrestrialization event and left the ancestor as aquatic, in the absence of additional evidence that favored one of the two options.

In addition, we are also getting results inconsistent with known information from the literature. MPR considered the ancestor of all arthropods and tardigrades to be terrestrial. This was changed to aquatic.

Gene repertoire evolutionary dynamics analyses

We used SonicParanoid2 v2.0.8 (fast mode)[207] to reconstruct orthologous groups (orthogroups) in our 973 species. A total of ~20M protein-coding sequences were clustered into 514,979 OGs, while ~5M were left unclustered. OGs were used as a proxy for gene families and used to infer gene gains and losses at different nodes using ETE v3.1.2[80]. For gains, we identified the most recent common ancestor (MRCA) of all species with genes in each OG to determine their age of origin (i.e. their phylostratigraphic origin but within the phylogenetic context of the dataset). For losses, we used a modification of the custom script used in[194] (see 'Code availability'). For each OG inferred to have originated at a given ancestral node, we assessed its presence or absence across descendant clades. A gene loss was inferred at a descendant node whenever all species descending from that node lacked genes in the OG. That is, the loss event is assigned to the MRCA of all species in which the OG was absent, provided that the OG was present in at least one species outside that clade within the parent lineage, and that the MRCA does not match the parental node. Gain-to-loss ratio was also computed at each node by dividing the number of OGs gained by the number of OGs lost.

BLASTp was run against the clustered nr database to get the identity of some OGs gained and lost. Full results are available in Figshare (see Data and Materials Availability section).

Functional enrichment analyses

We downloaded FANTASIA v1[208] GO functional predictions, based on similarity of embeddings from protein language models, from MATEdb2[79] or ran new inferences for the newly IsoSeq sequenced species.

To identify the function of OGs that were gained or lost at the origin of terrestrialisation lineages (containing at least two terrestrial species), we performed a GO enrichment of the proteins contained within those OGs from selected species. For each node, we selected the smallest number of species necessary to ensure complete OG representation. For gene gains, species were selected from within the clade defined by the node. For gene losses, species were selected from the sister lineage to the clade of interest (e.g., remipedes for Hexapoda). In Myriapoda and Onychophora, where including all species was unfeasible, we ranked species

by the number of OGs they contained, and selected the top 25. This subset recovered over 50% of the OGs, as shown by the cumulative increase in OGs with each added species (Suppl. Fig. S93c,d). For each terrestrialisation node, we performed separate GO enrichment analyses with topGO[209] for gene gains and losses. In both cases, the GO enrichment analyses were first done per species using the full proteome of that species as the background. We then combined the p-values using the ordmeta method [210] to identify enriched GO terms for genes gained or lost at the terrestrialisation node. Functional convergence was also assessed as explained below (see 'Functional convergence' section).

Stress experiments and RNA sequencing, and preprocessing

To pinpoint the gene repertoire composing the terrestrialisation genetic toolkit, we selected 17 aquatic and terrestrial species spanning 7 animal phyla (Nematoda, Onychophora, Arthropoda, Platyhelminthes, Mollusca, Annelida, Nemerta) (Fig. 1b, Table S71). We collected and used between 19 and 68 specimens per species to carry out the abiotic stress experiments detailed below (excepting in nematodes, where several specimens were pooled to do the experiments due to their microscopic size) (Tables S3-S19). Animals of similar size were included in each experiment to reduce age-related effects, and individuals were randomly assigned to treatment and control groups. The species included here were selected based on contrasting habitats (aquatic and terrestrial) within each phylum, allowing us to partially control for phylogenetic background while assessing responses to stress. In some phyla (molluscs and annelids) both freshwater and marine representatives were included within each phylum in addition to terrestrial ones, to capture broader taxonomic diversity. Chordata and Tardigrada were excluded due to limitations in experimental constraints (chordates) and specimen availability (tardigrades). The intensity and duration of each stressor were calibrated to elicit a genetic response without causing mortality, ensuring that the conditions fell within a biologically relevant range to reveal the gene repertoire underlying abiotic stress adaptation.

Natural light – Terrestrial species are regularly exposed to visible light as part of their daily environmental conditions, while many aquatic invertebrates typically experience limited or no direct light exposure. This ecological contrast suggests that light may act as a more frequent and biologically relevant stressor in terrestrial environments. To assess differential responses to visible light across aquatic and terrestrial species, we exposed individuals from all taxa to

ambient natural light for 15 minutes prior to dissection and flash freezing (experiment code: VL). This treatment enabled the identification of light-induced transcriptional responses and the evaluation of lineage-specific gene expression patterns potentially associated with adaptation to terrestrial environments.

Protection against UV radiation – Ultraviolet (UV-B) radiation (280–315 nm) is a potent genotoxic stressor that induces DNA damage primarily through the formation of cyclobutane pyrimidine dimers and 6–4 photoproducts[211]. These lesions cause structural distortions in the DNA helix, disrupting both transcription and replication. Two of the major repair pathways animals have evolved to mitigate such damage are: photoreactivation (light-dependent), which directly reverses DNA lesions using photolyase enzymes[212], and nucleotide excision repair (NER, or "dark repair"), a light-independent mechanism that removes damaged nucleotides through excision and resynthesis[213]. To investigate UV-induced stress responses and the dynamics of DNA repair, specimens were exposed to UV-B radiation (302 nm) at 8.48 W/m² using a UVM-28 EL Series UV lamp (Analytik Jena US) for durations ranging from 30 seconds to 3 minutes, depending on species-specific tolerance (see Table S72). Following exposure, animals were either flash-frozen immediately (experiment code: no recovery, UV0), or allowed to recover for 15 minutes or 24 hours under either visible light (experiment codes: UV15L, UV24L) or darkness (experiment codes: UV15D, UV24D). This design enabled us to distinguish between immediate damage responses and downstream repair mechanisms, and to experimentally separate light-dependent and light-independent pathways of DNA repair.

Osmoregulation experiments – Successful transition from marine to terrestrial environments requires major physiological adaptations in water balance regulation. Among the most critical challenges for terrestrial animals is the risk of dehydration due to increased evaporative water loss in air[214]. To assess desiccation tolerance in terrestrial species, we reduced ambient relative humidity to 29–30% RH and placed individuals in petri dishes where surface moisture was removed using filter paper, a fine brush, or silica gel beads, depending on the species' morphology. Behavioural and morphological indicators of dehydration, such as reduced activity or loss of cuticular shine, were carefully monitored, and the experiment was terminated once early signs of desiccation became apparent. In aquatic species, we examined osmoregulatory capacity by exposing marine animals to freshwater and freshwater animals to seawater for 30 seconds to 2 minutes, depending on species-specific tolerance (see Table S72 for conditions). These treatments simulate abrupt osmotic shifts and allow us to characterise

gene expression responses associated with maintaining ionic and water homeostasis—key processes likely to be involved in the evolution of terrestrial life (experiment code: OR).

Oxygen pressure experiments – One of the major physiological barriers to terrestrialisation is the shift from extracting oxygen in water to breathing air, which involves coping with different oxygen solubility and availability. To simulate this transition, we exposed animals to controlled oxygen extremes that reflect the lowest and highest oxygen concentrations known to support life in aquatic and terrestrial environments[215,216]. Hypoxic conditions (experiment code: Hypo) were generated using a HypoxyLab™ Hypoxia chamber (Oxford Optronix), except for Onychophora, where helium gas displacement was used as an alternative method due to logistic constraints during the field expedition. For hyperoxia (experiment code: Hyper), pure oxygen (ClearO2, UK) was sprayed into airtight glass containers until the target oxygen concentration was reached. Oxygen levels were continuously monitored using an OXY-1 SMA Single Channel Fiber Optic Oxygen Transmitter (PreSens Precision Sensing GmbH) and Planar Oxygen-Sensitive Spots affixed inside the experimental chambers. In hypoxia experiments, terrestrial animals were placed directly into airtight glass containers, which were then inserted into the HypoxyLab to achieve oxygen concentrations between 13–15% O_2. For aquatic species, the containers were pre-filled with seawater or freshwater (depending on the species), and the oxygen concentration was reduced to 8% O_2 before animals were introduced. For hyperoxia, terrestrial animals were transferred into pre-equilibrated airtight chambers after spraying with pure oxygen to achieve 40–42% O_2. Oxygen levels were maintained by additional injections as needed. In aquatic species, water was pre-oxygenated to reach 42–45% O_2 before animals were added. All exposures were conducted for 20 minutes to ensure sufficient physiological engagement without causing harm.

Reception of chemical cues – The transition from aquatic to terrestrial environments necessitated profound changes in sensory systems, particularly in the detection and processing of chemical cues. While chemosensation is a universal mechanism across animals, the nature of chemical stimuli differs markedly between water and air, requiring lineage-specific adaptations in receptor types and signaling pathways. To investigate how aquatic and terrestrial invertebrates respond to ecologically relevant chemosensory cues, we designed experiments using both attractive and aversive stimuli. For attractive cues, individuals were exposed to food sources representative of their natural diet (experiment code: CF, chemosensory food) (see Table S72 for species-specific details). To simulate warning signals, we used

decomposition-associated fluids. In brief, one conspecific individual per species was euthanised and allowed to decompose in phosphate-buffered saline (PBS), freshwater, or seawater, depending on the organism's native environment. The resulting fluid (dead conspecific) was used as an aversive chemosensory stimulus (experiment code: CD, chemosensory death). During trials, live specimens were placed individually in petri dishes (either dry or containing a small volume of water) while the stimulus (food or decomposed fluid) was introduced at the center. Behavioral responses, such as approach or avoidance, were monitored in real time. Upon a clear reaction (e.g., attraction or withdrawal), the experiment was concluded, and the sample was flash-frozen for RNA extraction.

For each treatment, we used several biological replicates per taxon and included a negative control (see Tables S3-S19 for detailed description). Specimens were flash frozen immediately after each experiment. In the case of onychophorans, samples were kept in RNAlater or TRIzol at -70ºC until processed. For small animals (for instance nematodes, nemerteans and molluscs), RNA was extracted from the whole specimen. For medium-sized specimens (as annelids or onychophorans), they were dissected into anterior and posterior parts (see Table S3-S19 for detailed description on replicates). RNA extractions after stress experiments were performed using the TRIzol® reagent (Invitrogen, USA) method following the manufacturer's instructions and using MaXtract® High Density tubes (Qiagen) to minimize DNA contamination prior to mechanical sample homogenization either by plastic rotor pestles or by ceramic mortar, depending on the sample. The concentration of all samples was assessed by Qubit RNA BR Assay kit (Thermo Fisher Scientific). Libraries were prepared with the TruSeq Stranded mRNA library preparation kit (Illumina), and sequenced on a NovaSeq 6000 (Illumina, 2 × 150 bp) for a minimum of 6Gb coverage. Summary statistics of sequenced data are shown in Tables S3-S19.

Because most animal species do not yet have sequenced genomes, and high-quality transcriptomes have been shown to sufficiently capture gene repertoire evolutionary dynamics, particularly when completeness metrics such as BUSCO scores are high[217], we generated full-length transcriptomes as reference datasets for co-expression network analysis. The same RNA extractions used for short-read sequencing were pooled together in each species to prepare an IsoSeq library. RNA samples were subjected to DNAse treatment using the Turbo DNA-free DNase (Invitrogen) following the manufacturer's instructions. SMRTbell libraries were generated following the procedure 'Preparing Iso-Seq® libraries using SMRTbell® prep

kit 3.0 (PN 102-396-000 REV02 APR2022)'. To enable pooling of multiple samples on 1 SMRTcell, libraries were made using barcoded adapters. The libraries were sequenced on a Sequel-IIe using Sequel II sequencing kit 2.0 and Binding kit 3.1 with 24 hr movie-time. Summary statistics of sequenced data is shown in Table S20. For the mollusc *Siphonaria pectinata*, due to the failure to successfully prepare an IsoSeq library, we built a reference transcriptome with Trinity v2.11.0[218] using 2 replicates per experiment condition and proceeded with the MATEdb2[79] pipeline.

Raw Illumina RNA-seq reads for all the experiments and species mentioned above were preprocessed using Trimmomatic v0.11.9[219] by removing adapters and ambiguous bases (MINLEN: 75, SLIDINGWINDOW: 4:15, LEADING: 10, TRAILING: 10, AVGQUAL: 30). Both raw and trimmed reads were assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) before any further analysis. For IsoSeq data, the data was demultiplexed with the IsoSeq pipeline v4.0.0 (https://github.com/PacificBiosciences/pbbioconda), and cDNA primers were removed using LIMA v2.7.1. PolyA tails and artificial concatemers were removed using isoseq refine v4.0.0, and the Full Length Non Concatemer (FLNC) reads were clustered into the final de-novo reference transcriptome using CD-HIT v4.8.1 (cd-hit-est, a=0.99)[220]. Each species-specific high-quality full-length reference transcriptome was evaluated using BUSCO (v5.4.7) using as reference database metazoa_odb10[221], with all reference transcriptomes showing very high BUSCO values (Table S20), hence being of enough quality to proceed with downstream analyses. Reference transcriptomes were translated into proteins using TransDecoder v5.5.0 (https://github.com/sghignone/TransDecoder), decontaminated using BlobTools v2 (https://blobtoolkit.genomehubs.org/blobtools2/), and the longest isoforms were obtained using a custom script for further functional analyses. The longest isoform proteomes were used as input in the orthology inference (see below).

Gene co-expression network construction

For each species, trimmed Illumina RNA-seq reads from each experiment were quasi-mapped to the corresponding species-specific IsoSeq transcriptome using Salmon v1.10.2[222]. Genes were filtered to retain only those with ≥10 counts in at least 90% of samples across all experiments, ensuring data sparsity did not confound network construction. Expression values were transformed using the regularised log (rlog) method implemented in DESeq2 v1.44.0[223],

which stabilizes variance across expression levels and is recommended for gene co-expression analysis. Batch effects were estimated and removed using the removeBatchEffect() function from limma v3.60.6[224], with automatic detection of hidden batch structure via surrogate variable analysis. Data integrity was assessed by identifying and excluding low-quality genes and outlier samples using the goodSamplesGenes() function from WGCNA v1.73[225]. Additional quality checks included principal component analysis and sample clustering via heatmaps (Fig. 276-292), performed using the plot_PCA() and plot_heatmaps() functions from BioNERO v1.12.0[226], prior to network construction.

Gene co-expression networks were independently constructed for each species with Weighted Gene Co-expression Network Analysis (WGCNA) using the reads generated by stress experiments. The soft-thresholding power β was determined using the pickSoftThreshold() function, with biweight midcorrelation (bicor) and a *signed hybrid* network type to fit the scale-free topology, at the threshold of $R^2 \geq 0.8$. The mean connectivity and scale-free topology fit were visualised using softConnectivity() and scaleFreePlot() to confirm the scale-free topology of the constructed networks. Networks were constructed using the blockwiseModules() function (*signed TOMType = signed, networkType = signed hybrid , deepSplit = 4, minModuleSize =30, mergeCutHeight = 0.15*), and module color assignments were extracted for downstream analysis.

To identify significant modules and extract hub genes, we first filtered co-expression modules based on their correlation with traits of interest (stress experiments), retaining only those with significant associations (p-value < 0.05). Hub genes were then extracted by selecting genes with both high gene significance (|GS| > 0.2) and high intramodular connectivity (|KME| > 0.8) within the significant modules and used for further analysis.

Proteomic validation of abiotic stress-induced genes

We performed untargeted LC-MS/MS proteomics to characterize proteins present in each species from those included in the abiotic stress experiments, using the control samples (see above). Samples were digested with TRIzol® reagent (Invitrogen, USA) and the protein phase was isolated, following the protocol in[227]. Samples were reduced with Tris (2-carboxyethyl) phosphine (100 mM, 60 ºC, 30 min) and alkylated in the dark with iodoacetamide (20 mM, 25 ºC, 30 min). The resulting protein extract was acidified with

phosphoric acid. Trypsin digestion and cleanup were performed according to the S-Trap manufacturer's instructions. Briefly, samples were loaded on S-Trap micro spin columns (ProtiFi, LLC cat # C02-micro) and washed extensively with 90% methanol containing 100 mM TEAB, pH 7.55. Trypsin was added directly to the microcolumn (1:10 w:w, 37ºC, 8h, Promega cat # V5113) and then peptides were eluted by serial addition of 50 mM TEAB, 0.1% formic acid, and 0.1% formic acid in 50% acetonitrile. Protein extracts were analysed on an Orbitrap Fusion Lumos mass spectrometer coupled to an EASY-nLC 1200 system. Peptides were separated via reverse-phase chromatography over a 50-cm C18 column, using a linear acetonitrile gradient. Data were acquired in data-dependent acquisition mode with dynamic exclusion, and spectra were searched using the Mascot engine v2.6 against species-specific de novo assembled transcriptomes.

Protein identification was based on high-confidence peptide matches with a false discovery rate (FDR) of ≤1% in Protein Discoverer v2.5. Only master proteins supported by unique peptides were retained, and the homogeneity of replicates was assessed by checking the percentage of proteins present in at least 2 replicates (Table S73). Quality control was ensured via regular BSA injections and QCloud performance monitoring[228]. Further experimental details follow those described in[227].

Phylostratigraphic analysis of module genes

To reconstruct the evolutionary history of stress-related genes, we mapped the experimentally identified, species-specific module-associated transcripts from each condition to their corresponding proteins in the reference proteomes and assigned them to orthogroups for phylostratigraphic analysis[48] using custom scripts (see 'Code availability'). Since species had a diverse number of module genes and these linked to varying numbers of orthogroups, we normalised the data by calculating percentages, i.e., dividing the number of genes assigned to each phylostratum (Node) by the total number of module genes per species. This approach ensured comparability across species and prevented bias due to differences in gene set sizes. The same approach was followed with module genes corroborated by proteomics. Results were visualised with iTOL v7[229]. RNA-seq data from *T. melanocephalum* and *T. longissimum* were merged at the genus level (*Tetrastemma*) because insufficient specimens were available from a single species to perform all stress experiments; as a result, different experiments were conducted on each of the two species.

The intersection of module-associated orthogroups between species and experiments was calculated and visualised with upset plots using the R package ComplexUpset v1.3.3[230] with 'exclusive_intersection' (i.e., distinct) mode.

Functional annotation of the 42 module-associated orthogroups shared across all species exposed to environmental stressors was performed with BLASTp against the NCBI non-redundant protein database and KEGG pathway annotation was performed with KEGG Mapper[231] and BlastKOALA[232]. The annotation was conducted using all the genes of *C. elegans* present in those 42 orthogroups as representative.

Functional convergence

To explore potential functional convergence between genes orchestrating the transcriptomic response to the different environmental stressors in each species, we employed two strategies.

The first strategy involved identifying metabolic pathways relevant for the response to each stress in each species. To achieve this, we assigned KEGG pathways using eggNOG-mapper v2.1.12[213] to each gene of the significant modules involved in the respective stress responses. Pathways categorised under 'Human Diseases' and 'Drug Development' in KEGG were excluded from further analysis. To account for differences in the number of stress-relevant genes identified per species, a pathway was considered relevant to the response if it comprised at least 1% of the total module genes for that specific species-stress context. Functional convergence was inferred by comparing the sets of relevant pathways across species and stress conditions. For visualisation, heatmaps were generated from pathway-percentage matrices, in which rows corresponded to KEGG pathways and columns to species. A summary heatmap combining all experimental conditions was constructed by retaining, for each pathway and species, the maximum observed percentage across conditions, and pathways present in more than 80% of species were retained. The same set of pathways was then used to generate condition-specific heatmaps. In all heatmaps, colour intensity represents the percentage of genes assigned to each pathway, with scales defined independently for each plot. The KEGG pathway "Metabolic pathways" was removed to avoid obscuring variation among more specific pathways. Additionally, unfiltered heatmaps were generated for each experimental condition to visualise the full pathway landscape. Module-level pathway composition was further calculated by quantifying the number of genes per module, the number of genes with

KEGG annotations, the number of unique pathways represented as well as the pathway richness. Pathway richness was defined as the number of unique KEGG pathways divided by the total number of genes in the module.

The second strategy involved generating "constellation plots" using the R package constellatoR v0.1.0) (https://github.com/MetazoaPhylogenomicsLab/constellatoR) from pre-computed "Biological Process" GO terms semantic similarity matrices. We created a custom script that used pygosemsim (https://github.com/mojaie/pygosemsim; same GO release as FANTASIA: March 22nd 2022) to estimate the Wang GO semantic similarity matrix between groups of GO terms using the Best-Match Average (BMA) method. Semantic similarity (simBMA) between 2 groups (g1 and g2) is calculated as follows:

$$sim_{BMA}(g1, g2) = \frac{\sum_{1=i}^{m} max_{1 \leq j \leq n} sim(go1i, go2j) + \sum_{1=j}^{n} max_{1 \leq j \leq m} sim(go1i, go2j)}{m+n}$$

where m and n are the number of GO terms annotated in g1 and g2, respectively; and go1i and go2j are the GO term in position i (from 1 to m) from the g1 and the GO term in position j (from 1 to n) from the g2, respectively. simBMASemantic similarity values range from 0 to 1, with 0 meaning no similarity and 1 meaning (almost) identical.

We calculated the functional similarity between different groups (in this case, we compared the functional similarity between (1) enriched GO terms for gained and lost orthogroups at independent terrestrialisation events, (2) species using the hub genes involved in the response to a stress condition in each species, and (3) OGs containing hub genes in the response to each stress condition).

These distance matrices were subsequently clustered in constellatoR using the Affinity Propagation cluster algorithm to identify groups of genes exhibiting high functional similarity, effectively delineating functional clusters, as previously described[18,50]. To visualize these results, we used scatter plots where the functional clusters are represented as polygons.

Machine learning to identify habitat-predictive orthogroups

To identify orthogroups that may have been parallely expanded, contracted, gained, or lost in terrestrial lineages, we trained a machine learning XGBoost algorithm. We used as input the Species-OGs counts table provided by SonicParanoid2 v2.0.3a5 (outgroups removed), where we used OGs as predictive features, and the habitat of each species as the target variable. To reduce the number of features to use in the training, we pre-filtered the table to remove groups of OGs that would not be informative to predict habitat: (1) species-specific multi-copy genes, (2) OGs exclusive from fully-aquatic phyla (e.g., Porifera or Priapulida). From a total of 514,979 OGs, we ended up with 443,457 OGs after both filters.

We evaluated the performance of the model using a K-fold stratified cross-validation. Stratification (Suppl. Fig. S272a) was performed based on "phylum" (Suppl. Fig. S272b) and habitat (Suppl. Fig. S272c) to account for sampling biases in the splitting of training and test sets due to overrepresentation of specific clades or habitats in our dataset ("effect of the phylogeny"). For "phylum", given that there were not enough species per phylum for the 5-fold cross-validation, we stratified based on the phyla that contain terrestrial and aquatic species, and for the rest, we grouped phyla into higher categories (i.e., no-Bilateria, Xenacoelomorpha, Deuterostomia, Ecdysozoa, and Lophotrochozoa).

As the number of features greatly exceeded the number of training examples in the dataset, even after pre-filtering uninformative OGs, we applied dynamic recursive feature elimination (dRFE) to iteratively reduce dimensionality based on feature importance. We used a modified implementation of Python functions from dRFEtools[234], incorporating a strategy that combines feature importances and model accuracies across folds, as described in the original algorithm[235].

In brief, the method begins by training the model on all features across 5 cross-validation folds. For each fold, features are ranked according to their importance in the model. These rankings are then aggregated across folds, weighted by the classification accuracy of each fold (see below), to produce an overall feature score. All features with an importance of 0 are removed in the first iteration. Then, for subsequent iterations, the bottom 30% of features (those with the lowest aggregated scores) are removed, and the process repeats with the remaining features. This continues until only one feature remains (Fig. S290d).

The ranking criterion of feature i-th is computed as follows[235]:

$$F_i^{rank} = \sum_{j=1}^{n} F_{i,j} \times \frac{(A_j^{learn} + A_j^{validation})}{|A_j^{learn} - A_j^{validation}| + \varepsilon}$$

where j=1,.., n is the number of cross-validation folds, $F_{i,j}$, A_j^{learn} and $A_j^{validation}$ are the feature importance, the learning accuracy, and the validation accuracy of j-th feature, respectively. ε is a real number with a very small value.

From all the iterations of the dRFE, we identified the minimum number of features needed to accurately predict the environment on our species (top 130 OGs) by checking the number of features used that maximised validation accuracy and minimised log-loss in the test set (Fig. 6b,c).

To characterize those top 130 OGs (Table S70), we performed BLASTp against the nr database, and used PANGEA[236] to do a KEGG and REACTOME pathway enrichment against *D. melanogaster, D. rerio,* and *M. musculus* references (using as input genes from those species contained in the top 130 OGs). For those OGs where there were no informative hits in any analyses, we checked the GO terms predicted by FANTASIA.

References

1. Selden, P. A. *Terrestrialization (Precambrian-Devonian) [Internet]. eLS*. (2005).
2. Vecoli, M., Clément, G. & Meyer-Berthaud, B. *The Terrestrialization Process: Modelling Complex Interactions at the Biosphere-Geosphere Interface*. (Geological Society of London, 2010).
3. Janvier, P. Terrestrialization: the early emergence of the concept. *Geol. Soc. Spec. Publ.* **339**, 5–9 (2010).
4. Garwood, R. J. & Edgecombe, G. D. Early terrestrial animals, evolution, and uncertainty. *Evolution (N. Y.)* **4**, 489–501 (2011).
5. Benton, M. J. Biodiversity on land and in the sea. *Geological Journal* **36**, 211–230

(2001).

6. Daley, A. C., Antcliffe, J. B., Drage, H. B. & Pates, S. Early fossil record of Euarthropoda and the Cambrian Explosion. *Proc Natl Acad Sci U S A* **115**, 5323–5331 (2018).

7. Buatois, L. A. *et al.* The Invasion of the Land in Deep Time: Integrating Paleozoic Records of Paleobiology, Ichnology, Sedimentology, and Geomorphology. *Integr Comp Biol* **62**, 297–331 (2022).

8. Cloudsley-Thompson, J. L. *Evolution and Adaptation of Terrestrial Arthropods*. (Springer Science & Business Media, 2012).

9. de Vries, J. & Archibald, J. M. Plant evolution: landmarks on the path to terrestrial life. *New Phytol* **217**, 1428–1434 (2018).

10. Dunlop, J. A., Scholtz, G. & Selden, P. A. Water-to-Land Transitions. *Arthropod Biology and Evolution* 417–439 (2013).

11. Van Straalen, N.M. Evolutionary terrestrialization scenarios for soil invertebrates. *Pedobiologia* **87-88**, 150753 (2021).

12. Allen, K. & Briggs, D. E. G. *Evolution and the Fossil Record*. (1989).

13. Long, J. A. & Gordon, M. S. The Greatest Step in Vertebrate History: A Paleobiological Review of the Fish‐Tetrapod Transition*. *Physiological and Biochemical Zoology* (2004) doi:10.1086/425183.

14. Benton, M. J. *Cowen's History of Life*. (John Wiley & Sons, 2019).

15. Minter, N. J. *et al.* The Prelude to Continental Invasion. *The Trace-Fossil Record of Major Evolutionary Events* 157–204 (2016).

16. Wang, K. *et al.* African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* **184**, 1362–1376.e18 (2021).

17. Meyer, A. *et al.* Giant lungfish genome elucidates the conquest of land by vertebrates.

Nature **590**, 284–289 (2021).

18. Aristide, L. & Fernández, R. Genomic insights into mollusk terrestrialization: Parallel and convergent gene family expansions as key facilitators in out-of-the-sea transitions. *Genome Biol. Evol.* **15**, (2023).
19. Benítez-Álvarez, L. *et al.* Genomic exaptation and regulatory landscape shifts as key mechanisms enabling flatworm terrestrialization. *Genome Biol. Evol.* **18**: evag063 (2026).
20. Schartl, M. *et al.* The genomes of all lungfish inform on genome expansion and tetrapod evolution. *Nature* **634**, 96–103 (2024).
21. Bi, X. *et al.* Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes. *Cell* **184**, 1377–1391.e14 (2021).
22. Martínez-Redondo, G. I. *et al.* Parallel duplication and loss of aquaporin-coding genes during the 'out of the sea' transition as potential key drivers of animal terrestrialization. *Mol. Ecol.* **32**, 2022–2040 (2023).
23. Fleming, J. F., Pisani, D. & Arakawa, K. The Evolution of Temperature and Desiccation-Related Protein Families in Tardigrada Reveals a Complex Acquisition of Extremotolerance. *Genome Biol Evol* **16**, (2024).
24. Little, C. *The Colonisation of Land: Origins and Adaptations of Terrestrial Animals*. (Cambridge University Press, 1983).
25. Willmer, P., Stone, G. & Johnston, I. *Environmental Physiology of Animals*. (John Wiley & Sons, 2009).
26. Cockell, C. S. Biological effects of high ultraviolet radiation on early earth--a theoretical evaluation. *J Theor Biol* **193**, 717–729 (1998).
27. Ache, B. W. & Young, J. M. Olfaction: diverse species, conserved principles. *Neuron* **48**, 417–430 (2005).

28. Benoit, J. B., McCluney, K. E., DeGennaro, M. J. & Dow, J. A. T. Dehydration Dynamics in Terrestrial Arthropods: From Water Sensing to Trophic Interactions. *Annu Rev Entomol* **68**, 129–149 (2023).

29. Minelli, A., Boxshall, G. & Fusco, G. *Arthropod Biology and Evolution: Molecules, Development, Morphology*. (Springer Science & Business Media, 2013).

30. Fernández, R. & Gabaldón, T. Gene gain and loss across the metazoan tree of life. *Nat. Ecol. Evol.* **4**, 524–533 (2020).

31. Richter, D. J., Fozouni, P., Eisen, M. B. & King, N. Gene family innovation, conservation and loss on the animal stem lineage. *Elife* **7**, e34226 (2018).

32. Domazet-Lošo, M., Široki, T., Šimičević, K. & Domazet-Lošo, T. Macroevolutionary dynamics of gene family gain and loss along multicellular eukaryotic lineages. *Nature Communications* **15**, 1–22 (2024).

33. Wolf, Y. I. & Koonin, E. V. Genome reduction as the dominant mode of evolution. *BioEssays* **35**, 829–837 (2013).

34. Joyce, W. *et al.* A Revised Perspective on the Evolution of Troponin I and Troponin T Gene Families in Vertebrates. *Genome Biol Evol* **15**, (2023).

35. Troyanovsky, B., Levchenko, T., Månsson, G., Matvijenko, O. & Holmgren, L. Angiomotin: an angiostatin binding protein that regulates endothelial cell migration and tube formation. *J Cell Biol* **152**, 1247–1254 (2001).

36. Leimeister, C., Externbrink, A., Klamt, B. & Gessler, M. Hey genes: a novel subfamily of hairy- and Enhancer of split related genes specifically expressed during mouse embryogenesis. *Mech Dev* **85**, 173–177 (1999).

37. Baba, N. & Sharma, H. M. Histochemistry of lactic dehydrogenase in heart and pectoralis muscles of rat. *J Cell Biol* **51**, 621–635 (1971).

38. Hua, X. *et al.* Multi-level transcriptome sequencing identifies COL1A1 as a candidate

marker in human heart failure progression. *BMC Medicine* **18**, 2 (2020).

39. Eghbali, M., Eghbali, M., Robinson, T. F., Seifter, S. & Blumenfeld, O. O. Collagen accumulation in heart ventricles as a function of growth and aging. *Cardiovasc Res* **23**, 723–729 (1989).
40. Brand, P. *et al.* The origin of the odorant receptor gene family in insects. *Elife* **7**, e38340 (2018).
41. Dadras, A. *et al.* Environmental gradients reveal stress hubs pre-dating plant terrestrialization. *Nat. Plants* **9**, 1419–1438 (2023).
42. Zhu, C. *et al.* WGCNA based identification of hub genes associated with cold response and development in Apis mellifera metamorphic pupae. *Front. Physiol.* **14**, 1169301 (2023).
43. Razalli, I. I. *et al.* Identification and validation of hub genes associated with biotic and abiotic stresses by modular gene co-expression analysis in Oryza sativa L. *Sci. Rep.* **15**, 8465 (2025).
44. Jeong, H., Mason, S. P., Barabási, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
45. Provero, P. Gene networks from DNA microarray data: centrality and lethality. *arXiv [cond-mat.stat-mech]* (2002) doi:10.48550/arXiv.cond-mat/0207345.
46. Carlson, M. R. J. *et al.* Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics* **7**, 40 (2006).
47. Serin, E. A. R., Nijveen, H., Hilhorst, H. W. M. & Ligterink, W. Learning from Co-expression Networks: Possibilities and Challenges. *Front Plant Sci* **7**, 444 (2016).
48. Domazet-Loso, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet* **23**, 533–539 (2007).

49. Bohutínská, M. & Peichel, C. L. Divergence time shapes gene reuse during repeated adaptation. *Trends Ecol Evol* **39**, 396–407 (2024).

50. Balart-García, P. *et al.* Parallel and convergent genomic changes underlie independent subterranean colonization across beetles. *Nat. Commun.* **14**, 3842 (2023).

51. Kültz, D. Molecular and evolutionary basis of the cellular stress response. *Annu Rev Physiol* **67**, 225–257 (2005).

52. López-Maury, L., Marguerat, S. & Bähler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics* **9**, 583–593 (2008).

53. Hatanaka R. *et al.* Signaling pathways in invertebrate immune and stress response. *Invertebrate Survival J.* **6**, 32–43 (2009).

54. Mao, L., Van Hemert, J. L., Dash, S. & Dickerson, J. A. Arabidopsis gene co-expression network and its functional modules. *BMC Bioinformatics* **10**, 346 (2009).

55. Ishikawa, A., Yamanouchi, S., Iwasaki, W. & Kitano, J. Convergent copy number increase of genes associated with freshwater colonization in fishes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20200509 (2022).

56. Davín, A. A. *et al.* A geological timescale for bacterial evolution and oxygen adaptation. *Science* **388**, eadp1853 (2025).

57. Gonçalves, C. *et al.* Diverse signatures of convergent evolution in cactus-associated yeasts. *PLOS Biology* **22**, e3002832 (2024).

58. Werck-Reichhart, D., Nelson, D. R. & Renault, H. Cytochromes P450 evolution in the plant terrestrialization context. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **379**, 20230363 (2024).

59. Loo, R. L., Chan, Q., Nicholson, J. K. & Holmes, E. Balancing the equation: A natural history of trimethylamine and trimethylamine-N-oxide. *J. Proteome Res.* **21**, 560–589

(2022).

60. Padilla, S., Prado, R. & Anitua, E. An evolutionary history of F12 gene: Emergence, loss, and vulnerability with the environment as a driver. *Bioessays* **45**, e2300077 (2023).

61. Green, J., Taylor, J. J., Hindes, A., Johnson, S. L. & Goldsmith, M. I. A gain of function mutation causing skeletal overgrowth in the rapunzel mutant. *Dev. Biol.* **334**, 224–234 (2009).

62. Lobanov, A. V., Hatfield, D. L. & Gladyshev, V. N. Eukaryotic selenoproteins and selenoproteomes. *Biochim. Biophys. Acta* **1790**, 1424–1428 (2009).

63. Lobanov, A. V. *et al.* Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life. *Genome Biol.* **8**, R198 (2007).

64. Hoitinga, P. K. & Birkeland, S. Pathway-level convergence: an underexplored aspect of convergent evolution. *Trends Genet* **41**, 853–867 (2025).

65. Doughty, T. W. *et al.* Stress-induced expression is enriched for evolutionarily young genes in diverse budding yeasts. *Nature Communications* **11**, 1–9 (2020).

66. Xie, K. T. *et al.* DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* (2019) doi:10.1126/science.aan1425.

67. Jacob, F. Evolution and Tinkering. *Science* (1977) doi:10.1126/science.860134.

68. Koonin, E. V. & Wolf, Y. I. Constraints and plasticity in genome and molecular-phenome evolution. *Nat Rev Genet* **11**, 487–498 (2010).

69. Watson-Zink, V. M., Grosberg, R. K., Lai, J. C. Y. & Bay, R. A. Transcriptomic responses of gecarcinid land crabs to acute and prolonged desiccation stress. *bioRxiv* 2024.07.03.601969 (2024) doi:10.1101/2024.07.03.601969.

70. Dong, Y., Krishnamoorthi, S., Tan, G. Z. H., Poh, Z. Y. & Urano, D. Co-option of plant gene regulatory network in nutrient responses during terrestrialization. *Nature Plants* **10**,

1955–1968 (2024).

71. Helsen, J. *et al.* Gene Loss Predictably Drives Evolutionary Adaptation. *Mol Biol Evol* **37**, 2989–3002 (2020).

72. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).

73. Xia, S., Chen, J., Arsala, D., Emerson, J. J. & Long, M. Functional innovation through new genes as a general evolutionary process. *Nature Genetics* **57**, 295–309 (2025).

74. Garwood, R. & Dunlop, J. Consensus and conflict in studies of chelicerate fossils and phylogeny. *Arachnologische Mitteilungen: Arachnology Letters* **66**:2-16 (2023).

75. Procheş, S., Polgar, G. & Marshall, D. J. K-Pg events facilitated lineage transitions between terrestrial and aquatic ecosystems. *Biol Lett* **10**, (2014).

76. Watson-Zink, V. M. Making the grade: Physiological adaptations to terrestrial environments in decapod crabs. *Arthropod Struct Dev* **64**, 101089 (2021).

77. Langschied, F. *et al.* Quest for Orthologs in the Era of Biodiversity Genomics. *Genome Biol Evol* **16**, (2024).

78. Liebeskind, B. J., McWhite, C. D. & Marcotte, E. M. Towards Consensus Gene Ages. *Genome Biol Evol* **8**, 1812–1823 (2016).

79. Martínez-Redondo, G. I. *et al.* MATEdb2, a collection of high-quality metazoan proteomes across the animal tree of life to speed up phylogenomic studies. *Genome Biol. Evol.* **16**, evae235 (2024).

80. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).

81. Ocaña-Pallarès, E. *et al.* Divergent genomic trajectories predate the origin of animals and fungi. *Nature* **609**, 747–753 (2022).

82. Laumer, C. E. *et al.* Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proc. Biol. Sci.* **286**, 20190831 (2019).

83. Kenny, N. J. *et al.* Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge Ephydatia muelleri. *Nature Communications* **11**, 1–11 (2020).

84. Plese, B. *et al.* Mitochondrial evolution in the Demospongiae (Porifera): Phylogeny, divergence time, and genome biology. *Molecular Phylogenetics and Evolution* **155**, 107011 (2021).

85. Simion, P. *et al.* A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Current Biology* **27**, 958–967 (2017).

86. DeBiasse, M. B. *et al.* A Cnidarian Phylogenomic Tree Fitted With Hundreds of 18S Leaves. *BSSB* **3**, (2024).

87. Kayal, E. *et al.* Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. *BMC Evolutionary Biology* **18**, 1–18 (2018).

88. Atherton, S. & Jondelius, U. Phylogenetic assessment and systematic revision of the acoel family Isodiametridae. *Zool J Linn Soc* **194**, 736–760 (2021).

89. Reich, A., Dunn, C., Akasaka, K. & Wessel, G. Phylogenomic analyses of Echinodermata support the sister groups of Asterozoa and Echinozoa. *PLoS One* **10**, e0119627 (2015).

90. Cannon, J. T. *et al.* Phylogenomic Resolution of the Hemichordate and Echinoderm Clade. *Current Biology* **24**, 2827–2832 (2014).

91. Zhang, Q.-L. *et al.* A Phylogenomic Framework and Divergence History of Cephalochordata Amphioxus. *Front Physiol* **9**, 1833 (2018).

92. Irisarri, I. *et al.* Phylotranscriptomic consolidation of the jawed vertebrate timetree. *Nature Ecology & Evolution* **1**, 1370–1378 (2017).

93. Yu, D. *et al.* Hagfish genome elucidates vertebrate whole-genome duplication events and

their evolutionary consequences. *Nature Ecology & Evolution* **8**, 519–535 (2024).

94. Hughes, L. C. *et al.* Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc Natl Acad Sci U S A* **115**, 6249–6254 (2018).
95. Bian, C. *et al.* Divergence, evolution and adaptation in ray-finned fish genomes. *Science China Life Sciences* **62**, 1003–1018 (2019).
96. Parey, E. *et al.* Genome structures resolve the early diversification of teleost fishes. *Science* **379**, 572–575 (2023).
97. Hime, P. M. *et al.* Phylogenomics Reveals Ancient Gene Tree Discordance in the Amphibian Tree of Life. *Syst Biol* **70**, 49–66 (2020).
98. Giribet, G. & Edgecombe, G. D. Current Understanding of Ecdysozoa and its Internal Phylogenetic Relationships. *Integr Comp Biol* **57**, 455–466 (2017).
99. Giacomelli, M. *et al.* CAT-Posterior Mean Site Frequencies Improves Phylogenetic Modeling Under Maximum Likelihood and Resolves Tardigrada as the Sister of Arthropoda Plus Onychophora. *Genome Biol Evol* **17**, evae273 (2024).
100. Herranz, M., Stiller, J., Worsaae, K. & Sørensen, M. V. Phylogenomic analyses of mud dragons (Kinorhyncha). *Molecular Phylogenetics and Evolution* **168**, 107375 (2022).
101. Ahmed, M. *et al.* Phylogenomic Analysis of the Phylum Nematoda: Conflicts and Congruences With Morphology, 18S rRNA, and Mitogenomes. *Front. Ecol. Evol.* **9**, (2022).
102. Bhat, A. H. *et al.* Morphological, Morphometrical and Molecular Characterization of Oscheius siddiqii Tabassum and Shahina, 2010 (Rhabditida, Rhabditidae) from India with Its Taxonomic Consequences for the Subgenus Oscheius Andrássy, 1976. *Biology* **10**, 1239 (2021).
103. Abolafia, J. & Vecchi, M. Redescription and phylogenetic analysis of the type species of

the genus Thorne, 1938 (Rhabditida, Panagrolaimidae), Thorne, 1938, including the first SEM study. *J Nematol* **53**, (2021).

104. Bleidorn, C., Schmidt-Rhaesa, A. & Garey, J. R. Systematic relationships of Nematomorpha based on molecular and morphological data. *Invertebrate Biology* **121**, 357–364 (2002).

105. Fleming, J. F. & Arakawa, K. Systematics of tardigrada: A reanalysis of tardigrade taxonomy with specific reference to Guil et al. (2019). *Zoologica Scripta* **50**, 376–382 (2021).

106. Guil, N. & Giribet, G. A comprehensive molecular phylogeny of tardigrades-adding genes and taxa to a poorly resolved phylum-level phylogeny. *Cladistics* **28**, 21–49 (2012).

107. Guil, N., Jørgensen, A. & Kristensen, R. An upgraded comprehensive multilocus phylogeny of the Tardigrada tree of life. *Zoologica Scripta* **48**, 120–137 (2019).

108. Nichols, P. B., Nelson, D. R. & Garey, J. R. A Family Level Analysis of Tardigrade Phylogeny. *Hydrobiologia* **558**, 53–60 (2006).

109. Stec, D., Vecchi, M., Calhim, S. & Michalczyk, Ł. New multilocus phylogeny reorganises the family Macrobiotidae (Eutardigrada) and unveils complex morphological evolution of the Macrobiotus hufelandi group. *Molecular Phylogenetics and Evolution* **160**, 106987 (2021).

110. Stec, D., Vecchi, M., Maciejowski, W. & Michalczyk, Ł. Resolving the systematics of Richtersiidae by multilocus phylogeny and an integrative redescription of the nominal species for the genus Crenubiotus (Tardigrada). *Scientific Reports* **10**, 1–20 (2020).

111. Baker, C. M., Buckman-Young, R. S., Costa, C. S. & Giribet, G. Phylogenomic Analysis of Velvet Worms (Onychophora) Uncovers an Evolutionary Radiation in the Neotropics. *Mol Biol Evol* **38**, 5391–5404 (2021).

112. Costa, C. S., Giribet FLS, G. & Pinto-Da-Rocha, R. Morphological and molecular phylogeny of Epiperipatus (Onychophora: Peripatidae): a combined approach. *Zool J Linn Soc* **192**, 763–793 (2020).

113. Giribet, G. & Edgecombe, G. D. The Phylogeny and Evolutionary History of Arthropods. *Current Biology* **29**, R592–R602 (2019).

114. Ballesteros, J. A. *et al.* Comprehensive Species Sampling and Sophisticated Algorithmic Approaches Refute the Monophyly of Arachnida. *Mol Biol Evol* **39**, msac021 (2022).

115. Xiong, Q. *et al.* Comparative Genomics Reveals Insights into the Divergent Evolution of Astigmatic Mites and Household Pest Adaptations. *Mol Biol Evol* **39**, msac097 (2022).

116. Pepato, A. R., Costa, S. G. dos S., Harvey, M. S. & Klimov, P. B. One-way ticket to the blue: A large-scale, dated phylogeny revealed asymmetric land-to-water transitions in acariform mites (Acari: Acariformes). *Molecular Phylogenetics and Evolution* **177**, 107626 (2022).

117. Ontano, A. Z. *et al.* Taxonomic Sampling and Rare Genomic Changes Overcome Long-Branch Attraction in the Phylogenetic Placement of Pseudoscorpions. *Mol Biol Evol* **38**, 2446–2467 (2021).

118. Kallal, R. J. *et al.* Converging on the orb: denser taxon sampling elucidates spider phylogeny and new analytical methods support repeated evolution of the orb web. *Cladistics* **37**, 298–316 (2021).

119. Wang, J., Bai, Y., Zhao, H., Mu, R. & Dong, Y. Reinvestigating the phylogeny of Myriapoda with more extensive taxon sampling and novel genetic perspective. *PeerJ* **9**, e12691 (2021).

120. Giribet, G. & Edgecombe, G. D. Conflict between datasets and phylogeny of centipedes: an analysis based on seven genes and morphology. *Proceedings of the Royal Society B: Biological Sciences* (2005) doi:10.1098/rspb.2005.3365.

121. Giribet, G. & Edgecombe, G. D. Stable phylogenetic patterns in scutigeromorph centipedes (Myriapoda : Chilopoda : Scutigeromorpha): dating the diversification of an ancient lineage of terrestrial arthropods. *Invert. Systematics* **27**, 485–501 (2013).

122. Bernot, J. P. *et al.* Major Revisions in Pancrustacean Phylogeny and Evidence of Sensitivity to Taxon Sampling. *Mol Biol Evol* **40**, msad175 (2023).

123. Schwentner, M., Richter, S., Christopher Rogers, D. & Giribet, G. Tetraconatan phylogeny with special focus on Malacostraca and Branchiopoda: highlighting the strength of taxon-specific matrices in phylogenomics. *Proceedings of the Royal Society B* doi:10.1098/rspb.2018.1524.

124. Oakley, T. H., Wolfe, J. M., Lindgren, A. R. & Zaharoff, A. K. Phylotranscriptomics to Bring the Understudied into the Fold: Monophyletic Ostracoda, Fossil Placement, and Pancrustacean Phylogeny. *Mol Biol Evol* **30**, 215–233 (2012).

125. Hiruta, S. F., Kobayashi, N., Katoh, T. & Kajihara, H. Molecular Phylogeny of Cypridoid Freshwater Ostracods (Crustacea: Ostracoda), Inferred from 18S and 28S rDNA Sequences. *jzoo* **33**, 179–185 (2016).

126. Luchetti, A. *et al.* Comparative genomics of tadpole shrimps (Crustacea, Branchiopoda, Notostraca): Dynamic genome evolution against the backdrop of morphological stasis. *Genomics* **113**, 4163–4172 (2021).

127. Mathers, T. C. *et al.* High lability of sexual system over 250 million years of evolution in morphologically conservative tadpole shrimps. *BMC Evolutionary Biology* **13**, 1–12 (2013).

128. Copilaş-Ciocianu, D., Borko, Š. & Fišer, C. The late blooming amphipods: Global change promoted post-Jurassic ecological radiation despite Palaeozoic origin. *Molecular Phylogenetics and Evolution* **143**, 106664 (2020).

129. Li, J.-Y., Zeng, C., Yan, G.-Y. & He, L.-S. Characterization of the mitochondrial genome

of an ancient amphipod Halice sp. MT-2017 (Pardaliscidae) from 10,908 m in the Mariana Trench. *Scientific Reports* **9**, 1–15 (2019).

130. Drumm, D. T. Phylogenetic Relationships of Tanaidacea (Eumalacostraca: Peracarida) Inferred from Three Molecular Loci. *J Crustacean Biol* **30**, 692–698 (2010).

131. Dimitriou, A. C., Taiti, S. & Sfenthourakis, S. Genetic evidence against monophyly of Oniscidea implies a need to revise scenarios for the origin of terrestrial isopods. *Scientific Reports* **9**, 1–10 (2019).

132. Lins, L. S. F., Ho, S. Y. W. & Lo, N. An evolutionary timescale for terrestrial isopods and a lack of molecular support for the monophyly of Oniscidea (Crustacea: Isopoda). *Organisms Diversity & Evolution* **17**, 813–820 (2017).

133. Wägele, J.-W., Holland, B., Dreyer, H. & Hackethal, B. Searching factors causing implausible non-monophyly: ssu rDNA phylogeny of Isopoda Asellota (Crustacea: Peracarida) and faster evolution in marine than in freshwater habitats. *Molecular Phylogenetics and Evolution* **28**, 536–551 (2003).

134. Javidkar, M., Cooper, S. J. B., King, R. A., Humphreys, W. F. & Austin, A. D. Molecular phylogenetic analyses reveal a new southern hemisphere oniscidean family (Crustacea : Isopoda) with a unique water transport system. *Invert. Systematics* **29**, 554–577 (2015).

135. Vereshchaka, A. L., Kulagin, D. N. & Lunina, A. A. A phylogenetic study of krill (Crustacea: Euphausiacea) reveals new taxa and co-evolution of morphological characters. *Cladistics* **35**, 150–172 (2019).

136. Wolfe, J. M. *et al.* A phylogenomic framework, evolutionary timeline and genomic resources for comparative studies of decapod crustaceans. *Proceedings of the Royal Society B* (2019) doi:10.1098/rspb.2019.0079.

137. Chak, S. T. C., Barden, P. & Baeza, J. A. The complete mitochondrial genome of the

eusocial sponge-dwelling snapping shrimp Synalpheus microneptunus. *Scientific Reports* **10**, 1–10 (2020).

138. Vereshchaka, A. L. The shrimp superfamily Sergestoidea: a global phylogeny with definition of new families and an assessment of the pathways into principal biotopes. *Royal Society Open Science* (2017) doi:10.1098/rsos.170221.

139. Bracken-Grissom, H. D. *et al.* A comprehensive and integrative reconstruction of evolutionary history for Anomura (Crustacea: Decapoda). *BMC Evolutionary Biology* **13**, 1–29 (2013).

140. Wolfe, J. M. *et al.* Convergent Adaptation of True Crabs (Decapoda: Brachyura) to a Gradient of Terrestrial Environments. *Syst Biol* **73**, 247–262 (2023).

141. Tsang, C. T. T., Schubart, C., Chu, K. H., Ng, P. K. L. & Tsang, L. M. Molecular phylogeny of Thoracotremata crabs (Decapoda, Brachyura): Toward adopting monophyletic superfamilies, invasion history into terrestrial habitats and multiple origins of symbiosis. *Molecular Phylogenetics and Evolution* **177**, 107596 (2022).

142. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* (2014) doi:10.1126/science.1257570.

143. Sun, X. *et al.* Phylomitogenomic analyses on collembolan higher taxa with enhanced taxon sampling and discussion on method selection. *PLOS ONE* **15**, e0230827 (2020).

144. Yu, D. *et al.* Phylogenomics of Elongate-Bodied Springtails Reveals Independent Transitions from Aboveground to Belowground Habitats in Deep Time. *Syst Biol* **71**, 1023–1031 (2022).

145. Yu, D. *et al.* Molecular phylogeny and trait evolution in an ancient terrestrial arthropod lineage: Systematic revision and implications for ecological divergence (Collembola, Tomocerinae). *Molecular Phylogenetics and Evolution* **154**, 106995 (2021).

146. Sendra, A., Jiménez-Valverde, A., Selfa, J. & Reboleira, A. S. P. Diversity, ecology,

distribution and biogeography of Diplura. *Insect Conservation and Diversity* **14**, 415–425 (2021).

147. Carapelli, A. *et al.* Going Deeper into High and Low Phylogenetic Relationships of Protura. *Genes* **10**, 292 (2019).
148. Wang, Z., Shi, Y., Qiu, Z., Che, Y. & Lo, N. Reconstructing the phylogeny of Blattodea: robust support for interfamilial relationships and major clades. *Scientific Reports* **7**, 1–8 (2017).
149. Johnson, K. P. *et al.* Phylogenomics and the evolution of hemipteroid insects. *Proceedings of the National Academy of Sciences* **115**, 12775–12780 (2018).
150. McKenna, D. D. *et al.* The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences* **116**, 24729–24737 (2019).
151. Kawahara, A. Y. *et al.* Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proceedings of the National Academy of Sciences* **116**, 22657–22663 (2019).
152. Peters, R. S. *et al.* Evolutionary History of the Hymenoptera. *Current Biology* **27**, 1013–1018 (2017).
153. Wiegmann, B. M. *et al.* Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences* **108**, 5690–5695 (2011).
154. Khalturin, K. *et al.* Polyzoa is back: The effect of complete gene sets on the placement of Ectoprocta and Entoprocta. *Sci Adv* **8**, eabo4400 (2022).
155. Lu, T.-M., Kanda, M., Satoh, N. & Furuya, H. The phylogenetic position of dicyemid mesozoans offers insights into spiralian evolution. *Zoological Lett* **3**, 6 (2017).
156. Marlétaz, F., Peijnenburg, K. T. C. A., Goto, T., Satoh, N. & Rokhsar, D. S. A New Spiralian Phylogeny Places the Enigmatic Arrow Worms among Gnathiferans. *Current Biology* **29**, 312–318.e3 (2019).

157. Gasmi, S. *et al.* Evolutionary history of Chaetognatha inferred from molecular and morphological data: a case study for body plan simplification. *Frontiers in Zoology* **11**, 1–25 (2014).

158. Melone, G., Ricci, C., Segers, H. & Wallace, R. L. Phylogenetic relationships of phylum Rotifera with emphasis on the families of Bdelloidea. *Hydrobiologia* **387**, 101–107 (1998).

159. Benítez-Álvarez, L. *et al.* Disentangling the evolutionary history of terrestrial planarians through phylogenomics. *bioRxiv* 2025.01.07.631701 (2025) doi:10.1101/2025.01.07.631701.

160. Laumer, C. E., Hejnol, A. & Giribet, G. Nuclear genomic signals of the 'microturbellarian' roots of platyhelminth evolutionary innovation. (2015) doi:10.7554/eLife.05503.

161. Orr, R. J. S. *et al.* Paleozoic origins of cheilostome bryozoans and their parental care inferred by a new genome-skimmed phylogeny. *Sci Adv* **8**, eabm7452 (2022).

162. Saadi, A. J. *et al.* Phylogenomics reveals deep relationships and diversification within phylactolaemate bryozoans. *Proceedings of the Royal Society B* (2022) doi:10.1098/rspb.2022.1504.

163. Fuchs, J., Iseto, T., Hirose, M., Sundberg, P. & Obst, M. The first internal molecular phylogeny of the animal phylum Entoprocta (Kamptozoa). *Mol Phylogenet Evol* **56**, 370–379 (2010).

164. Chen, Z. *et al.* A genome-based phylogeny for Mollusca is concordant with fossils and morphology. *Science* (2025) doi:10.1126/science.ads0215.

165. Song, H. *et al.* Scaphopoda is the sister taxon to Bivalvia: Evidence of ancient incomplete lineage sorting. *Proceedings of the National Academy of Sciences* **120**, e2302361120 (2023).

166. Irisarri, I., Uribe, J. E., Eernisse, D. J. & Zardoya, R. A mitogenomic phylogeny of chitons (Mollusca: Polyplacophora). *BMC Evolutionary Biology* **20**, 1–15 (2020).

167. Kocot, K. M., Todt, C., Mikkelsen, N. T. & Halanych, K. M. Phylogenomics of Aplacophora (Mollusca, Aculifera) and a solenogaster without a foot. *Proceedings of the Royal Society B* (2019) doi:10.1098/rspb.2019.0115.

168. Anderson, F. E. & Lindgren, A. R. Phylogenomic analyses recover a clade of large-bodied decapodiform cephalopods. *Molecular Phylogenetics and Evolution* **156**, 107038 (2021).

169. Lindgren, A. R. Molecular inference of phylogenetic relationships among Decapodiformes (Mollusca: Cephalopoda) with special focus on the squid Order Oegopsida. *Molecular Phylogenetics and Evolution* **56**, 77–90 (2010).

170. Uribe, J. E. & Zardoya, R. Revisiting the phylogeny of Cephalopoda using complete mitochondrial genomes. *J Molluscan Stud* **83**, 133–144 (2017).

171. Uribe, J. E. *et al.* A Phylogenomic Backbone for Gastropod Molluscs. *Syst Biol* **71**, 1271–1280 (2022).

172. Zhong, Z. *et al.* New mitogenomes in deep-water endemic Cocculinida and Neomphalida shed light on lineage-specific gene orders in major gastropod clades. *Front. Ecol. Evol.* **10**, 973485 (2022).

173. Cunha, T. J., Reimer, J. D. & Giribet, G. Investigating Sources of Conflict in Deep Phylogenomics of Vetigastropod Snails. *Syst Biol* **71**, 1009–1022 (2021).

174. Xu, T., Qi, L., Kong, L. & Li, Q. Mitogenomics reveals phylogenetic relationships of Patellogastropoda (Mollusca, Gastropoda) and dynamic gene rearrangements. *Zoologica Scripta* **51**, 147–160 (2022).

175. Cunha, T. J. & Giribet, G. A congruent topology for deep gastropod relationships. *Proceedings of the Royal Society B* (2019) doi:10.1098/rspb.2018.2776.

176. Uribe, J. E., Colgan, D., Castro, L. R., Kano, Y. & Zardoya, R. Phylogenetic relationships among superfamilies of Neritimorpha (Mollusca: Gastropoda). *Molecular Phylogenetics and Evolution* **104**, 21–31 (2016).

177. Yang, H. *et al.* Comparative Characterization of the Complete Mitochondrial Genomes of the Three Apple Snails (Gastropoda: Ampullariidae) and the Phylogenetic Analyses. *International Journal of Molecular Sciences* **19**, 3646 (2018).

178. Colgan, D. J., Ponder, W. F., Beacham, E. & Macaranas, J. Molecular phylogenetics of Caenogastropoda (Gastropoda: Mollusca). *Molecular Phylogenetics and Evolution* **42**, 717–737 (2007).

179. Goodheart, J. A. & Wägele, H. Phylogenomic analysis and morphological data suggest left-right swimming behavior evolved prior to the origin of the pelagic Phylliroidae (Gastropoda: Nudibranchia). *Organisms Diversity & Evolution* **20**, 657–667 (2020).

180. Krug, P. J. *et al.* Phylogenomic resolution of the root of Panpulmonata, a hyperdiverse radiation of gastropods: new insight into the evolution of air breathing. *Proceedings of the Royal Society B* (2022) doi:10.1098/rspb.2021.1855.

181. Razkin, O. *et al.* Molecular phylogeny of the western Palaearctic Helicoidea (Gastropoda, Stylommatophora). *Molecular Phylogenetics and Evolution* **83**, 99–117 (2015).

182. Doğan, Ö., Schrödl, M. & Chen, Z. The complete mitogenome of Arion vulgaris Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): mitochondrial genome architecture, evolution and phylogenetic considerations within Stylommatophora. *PeerJ* **8**, e8603 (2020).

183. González, V. L. *et al.* A phylogenetic backbone for Bivalvia: an RNA-seq approach. *Proceedings of the Royal Society B: Biological Sciences* (2015) doi:10.1098/rspb.2014.2332.

184. Smith, C. H. A High-Quality Reference Genome for a Parasitic Bivalve with Doubly Uniparental Inheritance (Bivalvia: Unionida). *Genome Biol Evol* **13**, evab029 (2021).

185. Lemer, S., González, V. L., Bieler, R. & Giribet, G. Cementing mussels to oysters in the pteriomorphian tree: a phylogenomic approach. *Proceedings of the Royal Society B: Biological Sciences* (2016) doi:10.1098/rspb.2016.0857.

186. Salvi, D. & Mariottini, P. Molecular taxonomy in 2D: a novel ITS2 rRNA sequence-structure approach guides the description of the oysters' subfamily Saccostreinae and the genus Magallana (Bivalvia: Ostreidae). *Zool J Linn Soc* **179**, 263–276 (2016).

187. Helm, C. *et al.* Convergent evolution of the ladder-like ventral nerve cord in Annelida. *Frontiers in Zoology* **15**, 1–17 (2018).

188. Li, Y. *et al.* Genomic adaptations to chemosymbiosis in the deep-sea seep-dwelling tubeworm Lamellibrachia luymesi. *BMC Biology* **17**, 1–14 (2019).

189. Xu, T. *et al.* The Morphology, Mitogenome, Phylogenetic Position, and Symbiotic Bacteria of a New Species of Sclerolinum (Annelida: Siboglinidae) in the South China Sea. *Front. Mar. Sci.* **8**, 793645 (2022).

190. Andrade, S. C. S. *et al.* Articulating 'Archiannelids': Phylogenomics and Annelid Relationships, with Emphasis on Meiofaunal Taxa. *Mol Biol Evol* **32**, 2860–2875 (2015).

191. Lemer, S. *et al.* Re-evaluating the phylogeny of Sipuncula through transcriptomics. *Molecular Phylogenetics and Evolution* **83**, 174–183 (2015).

192. Struck, T. H. *et al.* The Evolution of Annelids Reveals Two Adaptive Routes to the Interstitial Realm. *Current Biology* **25**, 1993–1999 (2015).

193. Tilic, E., Stiller, J., Campos, E., Pleijel, F. & Rouse, G. W. Phylogenomics resolves ambiguous relationships within Aciculata (Errantia, Annelida). *Molecular Phylogenetics*

and Evolution **166**, 107339 (2022).

194. Martín-Durán, J. M. *et al.* Conservative route to genome compaction in a miniature annelid. *Nat. Ecol. Evol.* **5**, 231–242 (2021).

195. Tilic, E., Sayyari, E., Stiller, J., Mirarab, S. & Rouse, G. W. More is needed—Thousands of loci are required to elucidate the relationships of the 'flowers of the sea' (Sabellida, Annelida). *Molecular Phylogenetics and Evolution* **151**, 106892 (2020).

196. Stiller, J., Tilic, E., Rousset, V., Pleijel, F. & Rouse, G. W. Spaghetti to a Tree: A Robust Phylogeny for Terebelliformia (Annelida) Based on Transcriptomes, Molecular and Morphological Data. *Biology* **9**, 73 (2020).

197. Erséus, C. *et al.* Phylogenomic analyses reveal a Palaeozoic radiation and support a freshwater origin for clitellate annelids. *Zool. Scr.* **49**, 614–640 (2020).

198. Tong, L. *et al.* The genome of medicinal leech (Whitmania pigra) and comparative genomic study for exploration of bioactive ingredients. *BMC Genomics* **23**, 1–13 (2022).

199. Andrade, S. C. S. *et al.* A Transcriptomic Approach to Ribbon Worm Systematics (Nemertea): Resolving the Pilidiophora Problem. *Mol Biol Evol* **31**, 3206–3215 (2014).

200. Chernyshev, A. V., Polyakova, N. E., Norenburg, J. L. & Kajihara, H. A molecular phylogeny of Tetrastemma and its allies (Nemertea, Monostilifera). *Zoologica Scripta* **50**, 824–836 (2021).

201. Chernyshev, A. V. & Polyakova, N. E. Distribution and Phylogenetic Position of the Antarctic Ribbon Worm Heteronemertes longifissa (Nemertea, Pilidiophora). *Water* **15**, 809 (2023).

202. Bapst, D. W., Schreiber, H. A. & Carlson, S. J. Combined Analysis of Extant Rhynchonellida (Brachiopoda) using Morphological and Molecular Data. *Syst Biol* **67**, 32–48 (2018).

203. Carlson, S. J. The evolution of brachiopoda. *Annu. Rev. Earth Planet. Sci.* **44**, 409–438

(2016).

204. Hirose, M., Fukiage, R., Katoh, T. & Kajihara, H. Description and molecular phylogeny of a new species of Phoronis (Phoronida) from Japan, with a redescription of topotypes of P. ijimai Oka, 1897. *Description and molecular phylogeny of a new species of Phoronis (Phoronida) from Japan, with a redescription of topotypes of P. ijimai Oka, 1897* **398**, 1–31 (2014).

205. WoRMS - World Register of Marine Species. https://www.marinespecies.org/aphia.php?p=popup&name=citation.

206. Bapst, D. W. paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution* **3**, 803–807 (2012).

207. Cosentino, S., Sriswasdi, S. & Iwasaki, W. SonicParanoid2: fast, accurate, and comprehensive orthology inference with machine learning and language models. *Genome Biol.* **25**, 195 (2024).

208. Martínez-Redondo, G. I. *et al.* Leveraging Natural Language Processing models to decode the dark proteome across the Animal Tree of Life. *Evolutionary Biology* (2024).

209. TopGO. *Bioconductor* https://bioconductor.org/packages/release/bioc/html/topGO.html.

210. Yoon, S., Baik, B., Park, T. & Nam, D. Powerful p-value combination methods to detect incomplete association. *Sci. Rep.* **11**, 6980 (2021).

211. Budden, T. & Bowden, N. A. The Role of Altered Nucleotide Excision Repair and UVB-Induced DNA Damage in Melanomagenesis. *International Journal of Molecular Sciences* **14**, 1132–1151 (2013).

212. Weber, S. Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1707**, 1–23 (2005).

213. Zachayus, A., Loup-Forest, J., Cura, V. & Poterszman, A. Nucleotide Excision Repair:

Insights into Canonical and Emerging Functions of the Transcription/DNA Repair Factor TFIIH. *Genes* **16**, 231 (2025).

214. Howard, J. M., Griffis, H.-B., Westendorf, R. & Williams, J. B. The influence of size and abiotic factors on cutaneous water loss. *Advances in Physiology Education* (2020) doi:10.1152/advan.00152.2019.

215. Verberk, W. C. E. P. *et al.* Does oxygen limit thermal tolerance in arthropods? A critical review of current evidence. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **192**, 64–78 (2016).

216. Mangum, C. & VAN Winkle, W. Responses of aquatic invertebrates to declining oxygen conditions. *Am. Zool.* **13**, 529–541 (1973).

217. Cheon, S., Zhang, J. & Park, C. Is phylotranscriptomics as reliable as phylogenomics? *Mol. Biol. Evol.* **37**, 3672–3683 (2020).

218. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652 (2011).

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

220. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

221. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).

222. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

223. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

224. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

225. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

226. Almeida-Silva, F. & Venancio, T. M. BioNERO: an all-in-one R/Bioconductor package for comprehensive and easy biological network reconstruction. *Funct. Integr. Genomics* **22**, 131–136 (2022).

227. García-Vernet, R. *et al.* A proteo-transcriptomic investigation of toxin evolution in planarians and their role in flatworm terrestrialization. *Evolutionary Biology* (2025).

228. Chiva, C. *et al.* QCloud: A cloud-based quality control system for mass spectrometry-based proteomics laboratories. *PLoS One* **13**, e0189209 (2018).

229. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).

230. Krassowski, M., Arts, M., Lagger, C. & Max. *Krassowski/complex-Upset: v1.3.5*. (Zenodo, 2022). doi:10.5281/ZENODO.3700590.

231. Kanehisa, M. & Sato, Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci* **29**, 28–35 (2020).

232. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**, 726–731 (2016).

233. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. EggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).

234. Benjamin, K. J. M., Katipalli, T. & Paquola, A. C. M. dRFEtools: dynamic recursive feature elimination for omics. *Bioinformatics* **39**, btad513 (2023).

235. Nguyen, H.-N. & Ohn, S.-Y. DRFE: Dynamic recursive feature elimination for gene identification based on random forest. in *Lecture Notes in Computer Science* 1–10 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2006).

236. Hu, Y. *et al.* PANGEA: a new gene set enrichment tool for Drosophila and common research organisms. *Nucleic Acids Res.* **51**, W419–W426 (2023).

Acknowledgements: We thank Luana Da Costa Monteiro, Nicholas Stroustrup and the Planarian Lab from Universitat de Barcelona for kindly providing nematode and planarian samples to initiate the cultures used in the stress experiments; Koryu Kin and Gonzalo Berceo for letting us use their HypoxiLab for the hypoxia experiments; Salvatore Cosentino for his assistance in the installation and bug fixing of SonicParanoid2; and Leandro Aristide and Lisandra Benítez-Álvarez for sharing with us their scripts for plotting the mirror trees. The CRG/UPF Proteomics Unit is part of the Spanish Infrastructure for Omics Technologies (ICTS OmicsTech). We also thank Centro de Supercomputación de Galicia (CESGA), particularly Pablo Rey for his assistance and guidance, and the HPC Drago from the Centro Superior de Investigaciones Científicas for access to computer resources. We declare the use of AI tools (ChatGPT, Perplexity) to polish grammar and clarity.

Funding:

Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya and ESF Investing in your future grant 2021 FI_B 00476 (GIMR)

Beatriu de Pinós fellowship from Secretaria d'Universitats i Recerca del Departament de Recerca i Universitats of the Generalitat de Catalunya Ref. BP 2021 00035 (FÁFÁ)

Ramón y Cajal fellowship Ref. RYC2023-043494-I (FÁFÁ)

MCIN/AEI /10.13039/501100011033 and FSE+ grant CEX2024-001494-S (FÁFÁ)

Spanish Ministry of Universities and the European Union Next Generation EU/PRTR Margarita Salas grant (RGV)

European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement no. 948281 (RF)
European Commission's Horizon Europe Research and Innovation programme grant agreement no. 101129751 OSCARS project (RF)
Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya AGAUR 2021-SGR00420 (RF)

Author contributions:

Conceptualization: GIMR, KE, RF

Specimens collection in fieldwork expeditions: FÁFÁ, KE, GIMR, JSO, IR, RGV, RF

Gene repertoire evolution analysis: GIMR

Transcriptomic and gene co-expression network analyses: KE

Stress experiments: KE, JSO, NE

Wet lab experiments: JSO, NE

Machine learning analysis: GIMR, AB

Proteomics analyses: JPG, RGV, LR, CC, ES, KE

Computational analyses assistance: BC, CVC

Interpretation and discussion: GIMR, KE, RF

Visualization: GIMR, KE, CVC, BC

Computational resources: AB, RF

Funding acquisition: RF

Supervision: RF

Writing – original draft: GIMR, KE, RF

Writing – review & editing: GIMR, KE, JSO, NE, FÁFÁ, BC, CVC, RGV, JPG, LR, IR, CC, ES, AB, RF

Competing interests:

Authors declare that they have no competing interests.

Data and Materials Availability:

RNA-seq and IsoSeq raw data are available at the SRA NCBI repository, under BioProject PRJEB111368. Mass spectrometry proteomics data is available in ProteomeXchange Consortium via PRIDE under accession number PXD064337. Output files for all main analyses are available at Figshare: https://figshare.com/s/90da0d7fbd3afdd6d5c0 (*10.6084/m9.figshare.29210009*). All code and customs scripts are available at the GitHub repository associated to this manuscript https://github.com/MetazoaPhylogenomicsLab/Martinez-Redondo-Eleftheriadi_et_al_2025_Genomic_basis_animal_terrestrialization (*final DOI to be provided*).

Tables S1 to S73 can be found in the following file: Supplementary Tables

Supplementary Materials

Headers for tables S1 to S73

Figs. S1 to S292

Supplementary Tables S1 to S73 can be found in the following file: Supplementary Tables

Table S1. Species used in the present study, including their species codes used in the analyses, taxonomic and habitat information, the proteome data source (i.e., transcriptome, genome, or IsoSeq), and whether it was used in stress experiments.

Table S2. OGs gained and lost, and gene-to-loss ratio at each terrestrialization event. Events defined only by one species are included, with the number of species-specific genes in parentheses. For these, gain-to-loss ratios are not calculated.

Table S3. Summary statistics of experimental RNA-seq data for *C. elegans* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S4. Summary statistics of experimental RNA-seq data for *L. marina* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S5. Summary statistics of experimental RNA-seq data for *P. aurorbis* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S6. Summary statistics of experimental RNA-seq data for *P. laevis* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S7. Summary statistics of experimental RNA-seq data for *L. oceanica* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S8. Summary statistics of experimental RNA-seq data for *O. nungara* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S9. Summary statistics of experimental RNA-seq data for *S. mediterranea* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S10. Summary statistics of experimental RNA-seq data for *P. turbinatus* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S11. Summary statistics of experimental RNA-seq data for *S. pectinata* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S12. Summary statistics of experimental RNA-seq data for *T. pisana* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S13. Summary statistics of experimental RNA-seq data for *P. acuta* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S14. Summary statistics of experimental RNA-seq data for *A. marina* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S15. Summary statistics of experimental RNA-seq data for *E. andrei* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S16. Summary statistics of experimental RNA-seq data for *H. medicinalis* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S17. Summary statistics of experimental RNA-seq data for *L. chalicophora* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S18. Summary statistics of experimental RNA-seq data for *T. melanocephalum* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S19. Summary statistics of experimental RNA-seq data for *T. longissimum* before and after preprocessing of reads, including the percentage of mapping to the de novo reference transcriptome.

Table S20. Summary statistics of experimental PacBio Iso-Seq data and de novo transcriptome assembly metrics for each species. Key statistics include the number of sequencing runs, total HiFi reads, yield, mean read length, read quality, number of transcripts after CD-HIT clustering, BUSCO completeness scores (metazoa_odb10), and number of predicted longest isoform proteins. One species (*S. pectinata*) was assembled using Illumina reads and Trinity due to lack of Iso-Seq data.

Table S21. Scale-free topology (SFT) analysis parameters for species-specific WGCNA.
The table includes the R^2 threshold used to assess scale-free topology fit, the soft-thresholding power (β) selected to approximate a scale-free network, and the resulting mean connectivity. In cases where no satisfactory SFT fit ($R^2 \geq 0.8$) was achieved, a default β value of 6 was applied as recommended by WGCNA guidelines.

Table S22. Statistically significant gene co-expression modules identified in *C. elegans* across experimental conditions.

Table S23. Statistically significant gene co-expression modules identified in *L. marina* across experimental conditions.

Table S24. Statistically significant gene co-expression modules identified in *P. aurorbis* across experimental conditions.

Table S25. Statistically significant gene co-expression modules identified in *P. laevis* across experimental conditions.

Table S26. Statistically significant gene co-expression modules identified in *L. oceanica* across experimental conditions.

Table S27. Statistically significant gene co-expression modules identified in *O. nungara* across experimental conditions.

Table S28. Statistically significant gene co-expression modules identified in *S. mediterranea* across experimental conditions.

Table S29. Statistically significant gene co-expression modules identified in *P. turbinatus* across experimental conditions.

Table S30. Statistically significant gene co-expression modules identified in *S. pectinata* across experimental conditions.

Table S31. Statistically significant gene co-expression modules identified in *T. pisana* across experimental conditions.

Table S32. Statistically significant gene co-expression modules identified in *P. acuta* across experimental conditions.

Table S33. Statistically significant gene co-expression modules identified in *A. marina* across experimental conditions.

Table S34. Statistically significant gene co-expression modules identified in *E. andrei* across experimental conditions.

Table S35. Statistically significant gene co-expression modules identified in *H. medicinalis* across experimental conditions.

Table S36. Statistically significant gene co-expression modules identified in *L. chalicophora* across experimental conditions.

Table S37. Statistically significant gene co-expression modules identified in *T. melanocephalum* across experimental conditions.

Table S38. Statistically significant gene co-expression modules identified in *T. longissimum* across experimental conditions.

Table S39. Hub genes from statistically significant WGCNA modules per experimental condition in *C. elegans*.

Table S40. Hub genes from statistically significant WGCNA modules per experimental condition in *L. marina*.

Table S41. Hub genes from statistically significant WGCNA modules per experimental condition in *P. aurorbis*.

Table S42. Hub genes from statistically significant WGCNA modules per experimental condition in *P. laevis*.

Table S43. Hub genes from statistically significant WGCNA modules per experimental condition in *L. oceanica*.

Table S44. Hub genes from statistically significant WGCNA modules per experimental condition in *O. nungara*.

Table S45. Hub genes from statistically significant WGCNA modules per experimental condition in *S. mediterranea*.

Table S46. Hub genes from statistically significant WGCNA modules per experimental condition in *P. turbinatus*.

Table S47. Hub genes from statistically significant WGCNA modules per experimental condition in *S. pectinata*.

Table S48. Hub genes from statistically significant WGCNA modules per experimental condition in *T. pisana*.

Table S49. Hub genes from statistically significant WGCNA modules per experimental condition in *P. acuta*.

Table S50. Hub genes from statistically significant WGCNA modules per experimental condition in *A. marina*.

Table S51. Hub genes from statistically significant WGCNA modules per experimental condition in *E. andrei*.

Table S52. Hub genes from statistically significant WGCNA modules per experimental condition in *H. medicinalis*.

Table S53. Hub genes from statistically significant WGCNA modules per experimental condition in *L. chalicophora*.

Table S54. Hub genes from statistically significant WGCNA modules per experimental condition in *T. melanocephalum*.

Table S55. Hub genes from statistically significant WGCNA modules per experimental condition in *T. longissimum*.

Table S56. Normalised percentages of module-associated OGs for each experiment per species

Table S57. Functional annotations of 42 module-associated orthogroups shared across all 17 species exposed to abiotic stressors using BLASTp and KEGG Mapper. Each row represents a gene within one of the 42 orthogroups. Protein functions were inferred via BLASTp searches against the NCBI non-redundant protein database and functionally

classified using KEGG Mapper. Annotations corresponding to uninformative hypothetical proteins or redundant entries were excluded.

Table S58. Summary of KEGG pathway composition across statistically significant modules in each species. Columns include the number of transcripts per module, number of translated proteins (longest isoforms), number of proteins annotated with KEGG pathways, number of unique pathways in each module, and pathway richness.

Table S59. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment CD. Habitat refers to species used in the experiments only; OGs may include additional species. "Aquatic" indicates clusters with hubs from aquatic species, "Terrestrial" from terrestrial species, and "Mixed" includes hubs from both.

Table S60. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment CF. Habitat refers to species used in the experiments only; OGs may include additional species. "Aquatic" indicates clusters with hubs from aquatic species, "Terrestrial" from terrestrial species, and "Mixed" includes hubs from both.

Table S61. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment Hyper. Habitat refers to species used in the experiments only; OGs may include additional species. "Aquatic" indicates clusters with hubs from aquatic species, "Terrestrial" from terrestrial species, and "Mixed" includes hubs from both.

Table S62. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment Hypo. Habitat refers to species used in the experiments only; OGs may include additional species. "Aquatic" indicates clusters with hubs from aquatic species, "Terrestrial" from terrestrial species, and "Mixed" includes hubs from both.

Table S63. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment OR. Habitat refers to species used in the experiments only; OGs may include additional species. "Aquatic" indicates clusters with hubs from aquatic species, "Terrestrial" from terrestrial species, and "Mixed" includes hubs from both.

Table S64. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment UV0. Habitat refers to species used in the experiments only; OGs may include additional species. "Aquatic" indicates clusters with hubs from aquatic species, "Terrestrial" from terrestrial species, and "Mixed" includes hubs from both.

Table S65. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment UV15D. Habitat refers to species used in the experiments only; OGs may include additional species. "Aquatic" indicates clusters with hubs from aquatic species, "Terrestrial" from terrestrial species, and "Mixed" includes hubs from both.

Table S66. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment UV15L. Habitat refers to species used in the experiments only;

OGs may include additional species. “Aquatic” indicates clusters with hubs from aquatic species, “Terrestrial” from terrestrial species, and “Mixed” includes hubs from both.

Table S67. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment UV24D. Habitat refers to species used in the experiments only; OGs may include additional species. “Aquatic” indicates clusters with hubs from aquatic species, “Terrestrial” from terrestrial species, and “Mixed” includes hubs from both.

Table S68. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment UV24L. Habitat refers to species used in the experiments only; OGs may include additional species. “Aquatic” indicates clusters with hubs from aquatic species, “Terrestrial” from terrestrial species, and “Mixed” includes hubs from both.

Table S69. Clusters of hub-containing orthogroups grouped by Gene Ontology terms using constellatoR for experiment VL. Habitat refers to species used in the experiments only; OGs may include additional species. “Aquatic” indicates clusters with hubs from aquatic species, “Terrestrial” from terrestrial species, and “Mixed” includes hubs from both.

Table S70. Top 130 OGs inferred by the Machine Learning model. Information on the model ranking, function (BLASTp, PANGEA pathways, and FANTASIA GO terms), number of species and sequences of OGs, their age of origin, together with stress-experiments information (whether hub genes are contained, how many, in which species, and which experiments).

Table S71. Species sampled for the study, including habitat type, sampling coordinates, locality information, and permitting status. Coordinates refer to new sample collection points.

Table S72. Species-specific experimental conditions.

Table S73. Summary of the proteomics statistics for each species. MP #1 Unique peptides indicates the number of master proteins with at least 1 Unique peptide. MP#1 in at least 2 replicates indicates the number of Master Proteins that are detected in at least 2 out of 3 replicates. MP#1 without contaminants indicates the final number of master proteins after removing the contaminants detected through LC-MS.

Supplementary Figures can be found in the following file: Supplementary_figures.pdf

Figure S1. Linear phylogenetic tree showing OG gains (in green) and losses (in magenta) across internal branches. The number of each node is shown in black. Terrestrial lineages are highlighted in yellow. Branches are coloured in blue for aquatic and brown for terrestrial habitat based on the ancestral habitat reconstruction.

Figure S2. Circular phylogenetic tree showing log10-transformed OG gain-to-loss ratios across internal branches in circles. Ratios higher than 1 are shown in green and represent a higher gain, while ratios between 0 and 1 represent higher OG losses and are shown in magenta. Protostomia node number (1002) is specified. Terrestrial lineages are highlighted in yellow. Branches are coloured in blue for aquatic and brown for terrestrial habitat based on the ancestral habitat reconstruction.

Figure S3. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Hexapoda.

Figure S4. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Hexapoda.

Figure S5. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Hexapoda.

Figure S6. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Hexapoda.

Figure S7. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Hexapoda.

Figure S8. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Hexapoda.

Figure S9. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Coenobitidae (Anomura).

Figure S10. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Coenobitidae (Anomura).

Figure S11. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Coenobitidae (Anomura).

Figure S12. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Coenobitidae (Anomura).

Figure S13. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Coenobitidae (Anomura).

Figure S14. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Coenobitidae (Anomura).

Figure S15. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Gecarcinidae (Brachyura).

Figure S16. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Gecarcinidae (Brachyura).

Figure S17. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Gecarcinidae (Brachyura).

Figure S18. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Gecarcinidae (Brachyura).

Figure S19. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Gecarcinidae (Brachyura).

Figure S20. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Gecarcinidae (Brachyura).

Figure S21. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Isopoda.

Figure S22. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Isopoda.

Figure S23. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Isopoda.

Figure S24. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Isopoda.

Figure S25. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Isopoda.

Figure S26. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Isopoda.

Figure S27. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Myriapoda.

Figure S28. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Myriapoda.

Figure S29. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Myriapoda.

Figure S30. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Myriapoda.

Figure S31. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Myriapoda.

Figure S32. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Myriapoda.

Figure S33. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Arachnida.

Figure S34. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Arachnida.

Figure S35. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Arachnida.

Figure S36. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Arachnida.

Figure S37. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Arachnida.

Figure S38. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Arachnida.

Figure S39. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Onychophora.

Figure S40. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Onychophora.

Figure S41. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Onychophora.

Figure S42. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Onychophora.

Figure S43. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Onychophora.

Figure S44. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Onychophora.

Figure S45. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Eutardigrada.

Figure S46. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Eutardigrada.

Figure S47. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Eutardigrada.

Figure S48. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Eutardigrada.

Figure S49. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Eutardigrada.

Figure S50. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Eutardigrada.

Figure S51. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Rhabditida.

Figure S52. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Rhabditida.

Figure S53. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Rhabditida.

Figure S54. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Rhabditida.

Figure S55. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Rhabditida.

Figure S56. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Rhabditida.

Figure S57. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Stylommatophora.

Figure S58. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Stylommatophora.

Figure S59. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Stylommatophora.

Figure S60. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Stylommatophora.

Figure S61. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Stylommatophora.

Figure S62. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Stylommatophora.

Figure S63. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Ellobiida.

Figure S64. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Ellobiida.

Figure S65. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Ellobiida.

Figure S66. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Ellobiida.

Figure S67. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Ellobiida.

Figure S68. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Ellobiida.

Figure S69. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Crassiclitellata.

Figure S70. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Crassiclitellata.

Figure S71. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Crassiclitellata.

Figure S72. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Crassiclitellata.

Figure S73. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Crassiclitellata.

Figure S74. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Crassiclitellata.

Figure S75. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Acteonemertidae.

Figure S76. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Acteonemertidae.

Figure S77. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Acteonemertidae.

Figure S78. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Acteonemertidae.

Figure S79. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Acteonemertidae.

Figure S80. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Acteonemertidae.

Figure S81. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Geoplanidae.

Figure S82. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Geoplanidae.

Figure S83. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Geoplanidae.

Figure S84. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Geoplanidae.

Figure S85. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Geoplanidae.

Figure S86. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Geoplanidae.

Figure S87. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs gained in Tetrapoda.

Figure S88. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs gained in Tetrapoda.

Figure S89. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs gained in Tetrapoda.

Figure S90. Treemap showing semantic similarity-based clustering of enriched Biological Process GO terms of OGs lost in Tetrapoda.

Figure S91. Treemap showing semantic similarity-based clustering of enriched Molecular Function GO terms of OGs lost in Tetrapoda.

Figure S92. Treemap showing semantic similarity-based clustering of enriched Cellular Component GO terms of OGs lost in Tetrapoda.

Figure S93. Gained and lost orthogroups at terrestrialisation nodes. a, Semantic similarity matrix of GO terms from orthogroups (OGs) gained. **b**, Semantic similarity matrix of GO terms from lost orthogroups. **c**, Cumulative count of orthogroup lost in Myriapoda as number of species increases. **d**, Cumulative count of OGs lost in Onychophora as number of species increases.

Figure S94. Soft-thresholding power selection for WGCNA of *C. elegans*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S95. Soft-thresholding power selection for WGCNA of *L. marina*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S96. Soft-thresholding power selection for WGCNA of *P. aurorbis*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S97. Soft-thresholding power selection for WGCNA of *P. laevis*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S98. Soft-thresholding power selection for WGCNA of *L. oceanica*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S99. Soft-thresholding power selection for WGCNA of *O. nungara*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S100. Soft-thresholding power selection for WGCNA of *S. mediterranea*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S101. Soft-thresholding power selection for WGCNA of *P. turbinatus*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S102. Soft-thresholding power selection for WGCNA of *S. pectinata*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S103. Soft-thresholding power selection for WGCNA of *T. pisana*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S104. Soft-thresholding power selection for WGCNA of *P. acuta*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S105. Soft-thresholding power selection for WGCNA of *A. marina*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S106. Soft-thresholding power selection for WGCNA of *E. andrei*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S107. Soft-thresholding power selection for WGCNA of *H. medicinalis*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S108. Soft-thresholding power selection for WGCNA of *L. chalicophora*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S109. Soft-thresholding power selection for WGCNA of *T. melanocephalum*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S110. Soft-thresholding power selection for WGCNA of *T. longissimum*.
The left panel shows the scale-free topology model fit (signed R^2) across soft-thresholding powers (β), and the right panel shows the mean connectivity at each β.

Figure S111. Module-trait correlations between gene co-expression modules of *C. elegans* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S112. Module-trait correlations between gene co-expression modules of *L. marina* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S113. Module-trait correlations between gene co-expression modules of *P. aurorbis* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S114. Module-trait correlations between gene co-expression modules of *P. laevis* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S115. Module-trait correlations between gene co-expression modules of *L. oceanica* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S116. Module-trait correlations between gene co-expression modules of *O. nungara* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S117. Module-trait correlations between gene co-expression modules of *S. mediterranea* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S118. Module-trait correlations between gene co-expression modules of *P. turbinatus* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S119. Module-trait correlations between gene co-expression modules of *S. pectinata* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S120. Module-trait correlations between gene co-expression modules of *T. pisana* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S121. Module-trait correlations between gene co-expression modules of *P. acuta* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S122. Module-trait correlations between gene co-expression modules of *A. marina* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S123. Module-trait correlations between gene co-expression modules of *E. andrei* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson

correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S124. Module-trait correlations between gene co-expression modules of *H. medicinalis* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S125. Module-trait correlations between gene co-expression modules of *L. chalicophora* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S126. Module-trait correlations between gene co-expression modules of *T. melanocephalum* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S127. Module-trait correlations between gene co-expression modules of *T. longissimum* and experimental stress conditions. Each row represents a module eigengene, and each column corresponds to a specific stress experiment. Numbers indicate Pearson correlation coefficients (r), with p-values in parentheses. Cell colors range from blue (negative correlation) to red (positive correlation).

Figure S128. Pre-metazoan evolutionary origins of hub-containing orthogroups and species-specific stress-response networks. a, Phylostratigraphic analysis showing that the majority of hub genes originate from old orthogroups, based on genes recovered through both transcriptomic and proteomic data. Colored squares in the tree indicate the branch leading to each node, using the same color scheme as in the accompanying barplot. Nodes from collapsed branches not shown in the tree are marked with black dots. Protostomia to Phylum and Phylum to Species segments indicate the sum of all the in-between branches. Grey bars indicate missing data for specific species. b, Upset plot showing overlap of hub-containing orthogroups resulted from species-specific co-expression networks. Dots connect species that have a number of hub gene orthogroups in common, the number indicated in each bar from the barplot on top. No orthogroup was shared among all terrestrial species, nor consistently within phyla, indicating lineage-specific gene recruitment.

Figure S129. Functional convergence of species-specific KEGG pathways across animal lineages under stress conditions. a, Heatmaps showing the proportion of KEGG pathways in gene sets under abiotic stress conditions**.** Each row represents a KEGG pathway, and each column corresponds to a species. The left heatmap displays a summary

heatmap including only pathways that are present in at least 80% of species, considering all experimental conditions combined. On the right, separate heatmaps are shown for each experimental condition, filtered by the same set of pathways (rows) as in the summary heatmap. On the right, separate heatmaps are shown for each experimental condition, filtered by the same set of pathways (rows) as in the summary heatmap. Numbers in parentheses next to species names indicate the total number of enriched pathways detected in that species for each condition. Color scales represent the percentage of genes in each pathway. **b**, Top 10 pathways per experiment.

Figure S130. Unfiltered pathway heatmap for experiment CD. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected pathways within CD and columns correspond to species. Color scales represent the percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S131. Unfiltered pathway heatmap for experiment CF. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected pathways within CF and columns correspond to species. Color scales represent the percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S132. Unfiltered pathway heatmap for experiment Hyper. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected pathways within Hyper and columns correspond to species. Color scales represent the percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S133. Unfiltered pathway heatmap for experiment Hypo. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected pathways within Hypo and columns correspond to species. Color scales represent the percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S134. Unfiltered pathway heatmap for experiment OR. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected

pathways within OR and columns correspond to species. Color scales represent the percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S135. Unfiltered pathway heatmap for experiment UV0. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected pathways within UV0 and columns correspond to species. Color scales represent the percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S136. Unfiltered pathway heatmap for experiment UV15D. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected pathways within UV15D and columns correspond to species. Color scales represent the percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S137. Unfiltered pathway heatmap for experiment UV15L. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected pathways within UV15L and columns correspond to species. Color scales represent the percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S138. Unfiltered pathway heatmap for experiment UV24D. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected pathways within UV24D and columns correspond to species. Color scales represent the percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S139. Unfiltered pathway heatmap for experiment UV24L. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected pathways within UV24L and columns correspond to species. Color scales represent the

percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S140. Unfiltered pathway heatmap for experiment VL. Heatmap shows the proportion of genes assigned to each KEGG pathway. Rows correspond to all detected pathways within VL and columns correspond to species. Color scales represent the percentage of genes assigned to each pathway within each species. Numbers in parentheses indicate the number of pathways exceeding the 1% threshold for each species in that experiment.

Figure S141. Functional convergence of species-specific, GO-annotated hub genes across animal lineages under the remaining stress conditions. Constellation plots of functions of hub genes across 17 aquatic and terrestrial species for each experimental condition. Each polygon represents a distinct functional cluster formed by species with similar GO annotations, despite no orthogroup-level overlap across species. Cluster exemplars indicate the species that act as centroid in each cluster. The presence of hub genes from multiple species within the same functional cluster demonstrates convergence in stress-response roles across independent lineages.

Figure S142. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 2.

Figure S143. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6.

Figure S144. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7.

Figure S145. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 9.

Figure S146. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 14.

Figure S147. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 23.

Figure S148. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 29.

Figure S149. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 37.

Figure S150. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 53.

Figure S151. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 63.

Figure S152. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 74.

Figure S153. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 96.

Figure S154. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 98.

Figure S155. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 114.

Figure S156. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 124.

Figure S157. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 153.

Figure S158. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 154.

Figure S159. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 155.

Figure S160. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 175.

Figure S161. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 193.

Figure S162. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 195.

Figure S163. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 228.

Figure S164. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 255.

Figure S165. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 257.

Figure S166. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 295.

Figure S167. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 342.

Figure S168. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 436.

Figure S169. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 459.

Figure S170. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 487.

Figure S171. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 492.

Figure S172. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 498.

Figure S173. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 706.

Figure S174. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 707.

Figure S175. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 737.

Figure S176. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 864.

Figure S177. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 896.

Figure S178. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 902.

Figure S179. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 912.

Figure S180. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 943.

Figure S181. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 977.

Figure S182. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 1022.

Figure S183. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 1202.

Figure S184. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 1243.

Figure S185. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 1484.

Figure S186. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 1589.

Figure S187. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 1647.

Figure S188. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 1761.

Figure S189. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 1810.

Figure S190. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 2014.

Figure S191. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 2088.

Figure S192. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 2153.

Figure S193. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 2188.

Figure S194. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 2529.

Figure S195. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 2725.

Figure S196. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 2729.

Figure S197. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 2783.

Figure S198. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 2977.

Figure S199. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 3068.

Figure S200. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 3070.

Figure S201. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 3192.

Figure S202. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 3227.

Figure S203. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 3244.

Figure S204. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 3340.

Figure S205. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 3377.

Figure S206. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 3725.

Figure S207. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 4119.

Figure S208. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 4721.

Figure S209. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 4723.

Figure S210. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 4797.

Figure S211. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 4937.

Figure S212. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 5092.

Figure S213. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 5724.

Figure S214. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 5743.

Figure S215. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 5822.

Figure S216. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 5845.

Figure S217. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 5881.

Figure S218. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6057.

Figure S219. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6081.

Figure S220. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6281.

Figure S221. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6359.

Figure S222. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6457.

Figure S223. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6604.

Figure S224. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6637.

Figure S225. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6693.

Figure S226. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6845.

Figure S227. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 6905.

Figure S228. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7048.

Figure S229. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7348.

Figure S230. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7361.

Figure S231. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7407.

Figure S232. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7504.

Figure S233. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7509.

Figure S234. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7533.

Figure S235. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7534.

Figure S236. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7566.

Figure S237. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 7796.

Figure S238. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 8030.

Figure S239. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 8323.

Figure S240. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 8389.

Figure S241. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 8407.

Figure S242. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 8721.

Figure S243. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 8831.

Figure S244. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 8943.

Figure S245. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 8985.

Figure S246. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 9574.

Figure S247. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 9700.

Figure S248. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 9701.

Figure S249. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 9705.

Figure S250. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 9719.

Figure S251. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 9991.

Figure S252. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10041.

Figure S253. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10090.

Figure S254. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10205.

Figure S255. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10298.

Figure S256. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10407.

Figure S257. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10427.

Figure S258. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10479.

Figure S259. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10510.

Figure S260. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10512.

Figure S261. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10656.

Figure S262. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10708.

Figure S263. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10713.

Figure S264. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10762.

Figure S265. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 10970.

Figure S266. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 11043.

Figure S267. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 11183.

Figure S268. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 11206.

Figure S269. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 11214.

Figure S270. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 11303.

Figure S271. Species tree showing the origin (orange dot), losses (pink dots), and gene count (black bars on the outer side) per species for OG 11582.

Figure S272. Machine learning data stratification, workflow, and results. a, Data splitting into 5 stratified folds based on habitat and phylogenetic categories. **b**, Proportion of each phylogenetic category ("phylum") in the testing and training datasets for each of the 5 folds. **c**, Proportion of each habitat in the testing and training datasets for each of the 5 folds. **d**, Dynamic recursive feature elimination (dRFE) workflow.

Figure S273. Full functions from the top 10 OGs as ranked by the machine learning model. The OG names were obtained with BLASTp. GO terms on the right were predicted using FANTASIA.

Figure S274. MPR ancestral habitat reconstruction (aquatic in blue, terrestrial in brown). Red branches represent ambiguous reconstruction.

Figure S275. Manually modified MPR ancestral habitat reconstruction based on known information from previous studies. Aquatic branches are in blue, terrestrial in brown.

Figure S276. PCA and heatmap of *C. elegans* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S277. PCA and heatmap of *L. marina* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S278. PCA and heatmap of *P. aurorbis* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S279. PCA and heatmap of *P. laevis* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S280. PCA and heatmap of *L. oceanica* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S281. PCA and heatmap of *O. nungara* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S282. PCA and heatmap of *S. mediterranea* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S283. PCA and heatmap of *P. turbinatus* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S284. PCA and heatmap of *S. pectinata* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S285. PCA and heatmap of *T. pisana* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S286. PCA and heatmap of *P. acuta* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S287. PCA and heatmap of *A. marina* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S288. PCA and heatmap of *E. andrei* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S289. PCA and heatmap of *H. medicinalis* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S290. PCA and heatmap of *L. chalicophora* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S291. PCA and heatmap of *T. melanocephalum* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).

Figure S292. PCA and heatmap of *T. longissimum* RNA-seq samples using normalized counts (top) and after batch effect correction (bottom).