EVOLUTIONARY BIOLOGY

Marine-montane transitions coupled with gill and genetic convergence in extant crustacean

Hongguang Liu[1,2]†, Yami Zheng[1,3]†, Bingyue Zhu[1,3], Yan Tong[1,3], Wenpei Xin[1], Han Yang[1,3], Pengyu Jin[1], Yueyao Hu[1,3], Mengyi Huang[1,3], Wanjin Chang[4], Francesco Ballarin[5], Shuqiang Li[1]*, Zhonge Hou[1]*

Marine-terrestrial transition represents an important aspect of organismal evolution that requires numerous morphological and genetic innovations and has been hypothesized to be caused by geological changes. We used talitrid crustaceans with marine-coastal-montane extant species at a global scale to investigate the marine origination and terrestrial adaptation. Using genomic data, we demonstrated that marine ancestors repeatedly colonized montane terrestrial habitats during the Oligocene to Miocene. Biological transitions were well correlated with plate collisions or volcanic island formation, and top-down cladogenesis was observed on the basis of a positive relationship between ancestral habitat elevation and divergence time for montane lineages. We detected convergent variations of convoluted gills and convergent evolution of *SMC3* associated with montane transitions. Moreover, using CRISPR-Cas9 mutagenesis, we proposed that *SMC3* potentially regulates the development of exites, such as talitrid gills. Our results provide a living model for understanding biological innovations and related genetic regulatory mechanisms associated with marine-terrestrial transitions.

INTRODUCTION

Transitions from marine to terrestrial, including montane life, have occurred multiple times in diverse organisms (e.g., vertebrates, insects, and angiosperms), and they represent one of the most important evolutionary changes (*1*–*5*). These biological shifts have been hypothesized to be caused by geological changes and require numerous morphological and genetic innovations (*1*–*4*). Notable examples include the water-to-land transition of tetrapods due to drastic continent moving and severe drought in the Devonian, coupled with the evolution of the lung respiratory system and related genetic regulatory networks during this transition (*1*–*3*, *6*).

Marine-montane transitions often occurred in the continental margins, where preexisting marine habitat transitioned into air-exposed landscape by a series of geological changes and created conditions favoring organisms that transitioned from the sea to land and mountains (*7*). During the Oligocene to Miocene, the global geological setting of sea and land was greatly modified by tectonic collisions (India/Eurasia plate and Australia/Sundaland plate), island/mountain emergence (the Indo-Australian Archipelago), and volcanic activities (Southeast Australia, Tasmania) (*8*). However, detailed spatiotemporal investigations on the marine-montane transition and the relationship to these geological changes remain scarce because the animals involved in the ancient transitions are extinct.

Talitrid crustaceans are found worldwide and have extant marine, coastal, and montane species (Fig. 1A). Therefore, they could be a living model for investigating the marine-terrestrial transition and related morphological and genetic innovations. The montane talitrids are truly terrestrial; they inhabit forest and exhibit a unique jumping behavior into air. By contrast, their marine sister groups swim in seawater. Their success in transitioning to montane habitats is thought to have resulted from invasion of intertidal habitats by marine ancestors along with the innovation of respiratory gills with convoluted lobes, folds, and filaments, which provide a larger surface for gas exchange (*9*). Their current distributions in the continental margins or island arcs and gill changes between habitats indicate that this group may bear a signature of geological changes and evolved morphological adaptation during the marine-montane transition.

In this study, we reconstructed a global spatiotemporal framework of talitrids based on de novo genome assembly and obtained genome-wide data of 851 individuals spanning marine, coastal, and montane habitats (Fig. 1, B and C). Geometric morphometric analyses of respiratory gill features were performed to explore phenotypic innovations in montane talitrids. To investigate the genomic basis of morphological innovation and convergent evolution during the marine-montane transition, we conducted genomic and transcriptomic comparisons, performed hypothermia/hypoxia acclimation experiments, and identified structural maintenance of chromosomes 3 (*SMC3*) as a functional gill-related gene. CRISPR-Cas9 knockout validated the potential role of *SMC3* in regulating the gill phenotype. In summary, we explored the marine-montane transition and provided genetic evidence for the morphogenesis of gill innovation in talitrids, which serve as a living model system for elucidating marine origination and montane adaptation.

RESULTS

Genome features and phylogenomic data assembly

To establish a solid spatiotemporal framework, we first generated a near-chromosomal-level genome of *Morinoia aosen* and five draft

[1]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. [2]Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Yezin, Nay Pyi Taw 05282, Myanmar. [3]University of Chinese Academy of Sciences, Beijing 100049, China. [4]Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, 43900 Sepang, Selangor, Malaysia. [5]Systematic Zoology Laboratory, Department of Biological Sciences, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji-shi, 192-0397, Tokyo, Japan.
*Corresponding author. Email: houze@ioz.ac.cn (Z.H.); lisq@ioz.ac.cn (S.L.)
†These authors contributed equally to this work.

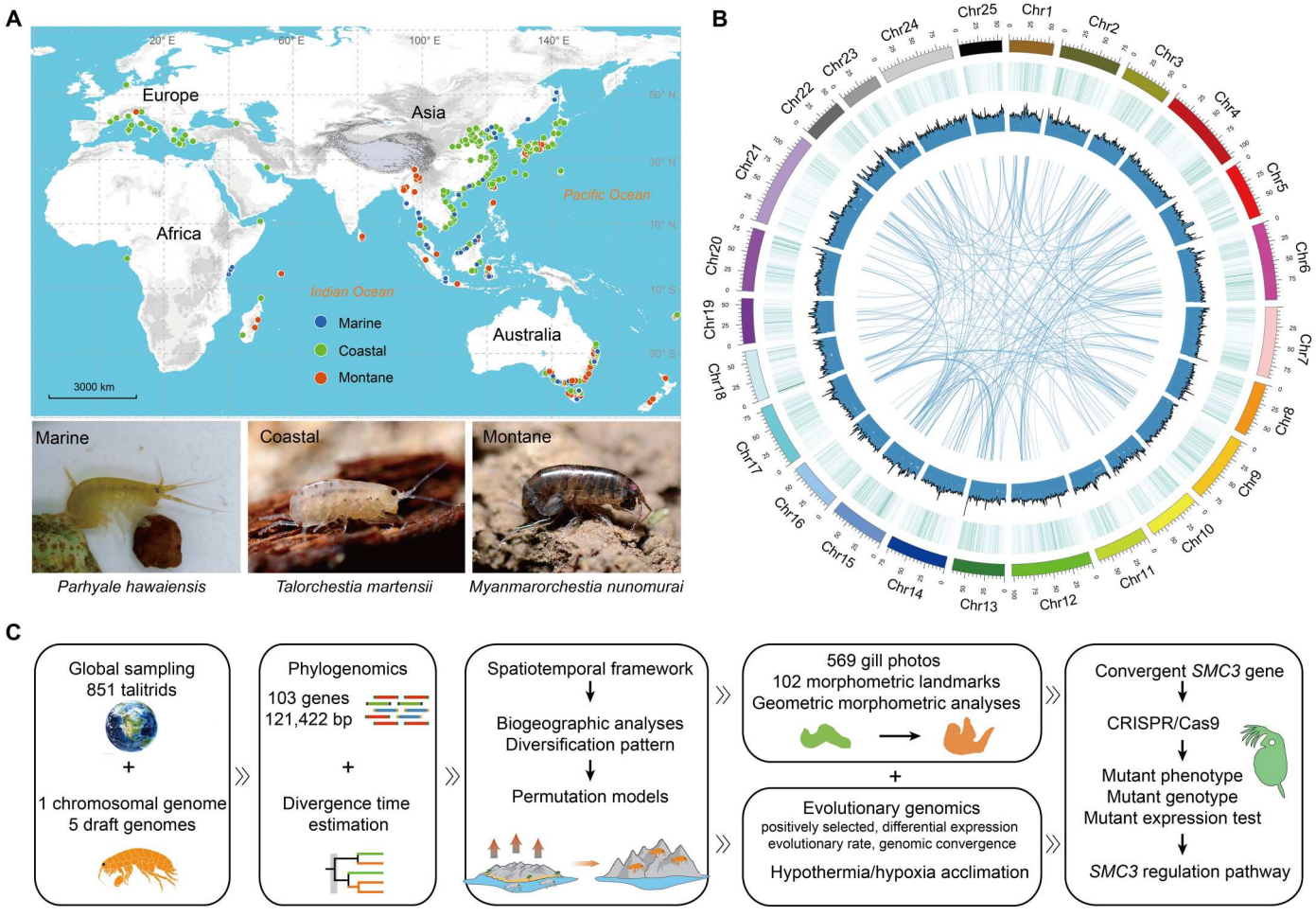

Fig. 1. Global sampling and chromosome-level genome assembly. (**A**) Collection sites included in this study. Colored circles on the map and photos below correspond to talitrids inhabiting different habitats. Photograph by T. Jiang and Z. Yao. (**B**) Circos plot of *M. aosen* genomic features. From the outer circle to the inner circle, four types of information (25 chromosomes, gene density, GC content, and synteny blocks between different chromosomes) are shown. The statistic was based on a 500-kb sliding window. (**C**) Workflow showing the dataset and main analyses used in this study.

genomes of talitrids. Then, we designed a genomic-scale dataset for Amphipoda according to exons from these genomes (Fig. 1C and tables S1 and S2).

We generated a 1.99-Gb contig-level assembly of *M. aosen* based on the approximately 60× coverage of PacBio long-read data and 121.84× coverage of short-read data; the contig number was 1783, and there was a high contig N50 value of 4.89 Mb. We obtained 135× coverage of Hi-C data and produced an approximately 1.99-Gb near-chromosomal-level genome. Approximately 99.96% of the draft genome sequences could be anchored, which resulted in 25 near-chromosomal scaffolds and four other scaffolds with a scaffold N50 value of 77.7 Mb (fig. S1 and tables S3 to S5). The Benchmarking Universal Single-Copy Orthologs (BUSCO) genome completeness score was 94.5% (table S6). Our *M. aosen* genome showed a higher scaffold N50 value and completeness assessment score among published crustacean genomes, which indicated that it is a high-quality reference genome. We also generated draft genome assemblies for five species; details are provided in tables S6 and S7.

On the basis of the high-quality genome and draft genomes, the designed phylogenomic dataset contained 96 loci from next-generation sequencing (NGS) data and seven loci from Sanger sequencing for 71 marine, 421 coastal, and 359 montane individuals (fig. S2 and table S1). The concatenated dataset contained 851 individuals with 121,422 nucleotides [NGS, 113,559 base pairs (bp); Sanger, 7863 bp]. On the basis of the molecular species delimitation using general mixed yule coalescent (GMYC) method (*10*), we composed a 397-tip dataset at the species level. Detailed information on samples, molecular operational taxonomic units (MOTUs), and sequence primers is presented in tables S1, S2, and S8.

Our innovative phylogenomic dataset was highly efficient for talitrid polymerase chain reaction (PCR) experiments because 80 markers showed more than 80% amplification (fig. S3, A and B). Compared with the varying percentages of variable sites of Sanger sequencing markers, such as cytochrome c oxidase subunit II (*COII*) with 34.01% and 18*S* ribosomal RNA with 6.42%, variable sites of the 96 nuclear markers were stable and ranged from 12.1 to 21.8%, with length ranging from 599 to 1634 bp (average, 1183 bp; fig. S3C).

The inferred maximum likelihood, Bayesian inference, and species tree phylogenies based on concatenation of the 851- and

397-tip datasets achieved identical and well-supported resolution of relationships among all major lineages (figs. S3D, S4, and S5). The multidimensional scaling plot from unweighted Robinson-Foulds distances of the individual gene tree and the concatenated tree indicate that the majority of the NGS markers overlapped with each other, except for three markers that had relatively low sequence recovery rates. Such results revealed that the 96 nuclear markers are stable and informative for phylogenetic reconstruction. Moreover, the NGS markers also showed a wide application for the order Amphipoda; seven Gammaridae samples received nearly 50% amplification (fig. S3, A and B).

Evolutionary history of marine-montane habitat transitions

Our phylogenetic reconstruction indicated that Talitridae formed a monophyletic group. The divergence time estimation and ancestral habitat reconstruction suggested the Talitridae originated in a marine habitat and then colonized the coast at 56.16 [95% highest posterior density (HPD): 70.46 to 48.72] million years (Ma) ago. It contains seven major clades (Fig. 2A, figs. S4 to S8, and tables S9 to S12).

The Southeast Asian mangrove clade, including the genus *Cochinorchestia*, is a basal group that inhabits mangrove forests along Southeast Asian and Australian coasts. The Myanmar-Australia montane clade diversified at 53.46 (95% HPD: 67.61 to 46.23) Ma ago and split into local geographic lineages in Myanmar at 26.6 (95% HPD: 32.04 to 23.07) Ma ago, South Australia at 20.37 (95% HPD: 23.25 to 17.64) Ma ago, Tasmania at 21.94 (95% HPD: 26.43 to 19.41) Ma ago, Southeast Australia at 19.47 (95% HPD: 21.67 to 16.92) Ma ago, and the Southern Alps of New Zealand at 18.6 (95% HPD: 23.36 to 15.11) Ma ago. The global *Talitroides* clade started diversifying at 36.56 (95% HPD: 51.29 to 27.94) Ma ago and was formed by terrestrial species from global mountains and some coastal species distributed along southeastern Australia. The Indo-West Pacific and Mediterranean *Talorchestia* clade contains coastal species belonging to genera *Notorchestia* and *Talorchestia*. The island montane clade consists of a coastal lineage from Australia and five montane lineages from different islands, including Luzon Island in the Philippines (25.56 Ma ago with 95% HPD: 32.12 to 21.41 Ma ago), the Japanese islands (24.52 Ma ago with 95% HPD: 29.06 to 20.94 Ma ago), Mount Kinabalu, Borneo (10.28 Ma ago with 95% HPD: 11.94 to 8.7 Ma ago), Northland New Zealand (30.65 Ma ago with 95% HPD: 38.35 to 25.94 Ma ago), and Sulawesi (14.72 Ma ago with 95% HPD: 16.4 to 12.84 Ma ago). The tropical *Floresorchestia* clade is distributed along the Indo-West Pacific coast, of which montane lineages are monophyletic originated at 24.46 (95% HPD: 27.12 to 21.55) Ma ago and are distributed in the Malay Peninsula, Myanmar, and southwest China. The West Pacific *Platorchestia* clade is widely distributed at the margin of East Asia and the Northwest Pacific coast; it is considered a coastal clade because the mean elevation of inland *Platorchestia* species is below 100 m and some of these species are distributed near the coast.

Biogeographic inference and ancestral habitat reconstruction indicated that talitrids independently colonized montane habitats at least four times during the Oligocene to Miocene (Fig. 2, A and B, and figs. S7 and S8). The Myanmar-Australia montane clade originated along a broad portion of the coast in the Indo-West Pacific region and then colonized montane habitats of continental margins. The island montane clade originated in coastal habitats that spanned the Japanese islands, Southeast Asia, and Australia, and then colonized montane habitats of the West Pacific islands. The Sundaland montane *Floresorchestia* is hypothesized to have originated along the Southeast Asia coast and then colonized Sundaland and Yunnan, China. *Talitroides* originated along the Australia coast; then, the terrestrial *Talitroides topitotum* became globally distributed by sylviculture (*9*).

Diversification dynamics and top-down cladogenesis of montane lineages

To investigate diversity patterns in each montane lineage and detect the biodiversity difference between the coastal and montane lineages, we selected eight montane lineages (Myanmar, Southeast Australia, South Australia, Tasmania, Japanese islands, Sulawesi, Philippines, and Northland New Zealand) and four coastal lineages (*Talorchestia*, *Platorchestia*, *Floresorchestia*, and *Notorchestia* lineages) based on the sampling completeness.

We first investigated the elevational diversification within each montane lineage. We inferred the ancestral habitat elevation for internal nodes of montane lineages and found that the elevation of the most recent common ancestor (MRCA) was higher than the average or median elevation of descendent species. Meanwhile, an overall trend of a positive relationship was detected between ancestral elevation and divergence time of internal nodes within each montane lineage (Fig. 2C and table S13). The nodes with old divergence times were associated with high-elevation habitat, and the nodes with young divergence times were associated with low-elevation habitat. These findings indicate that the ancestor of montane talitrids colonized higher elevations and then diversified downslope. We defined this scenario as top-down cladogenesis.

Elevational gradients of mountains lead to greater ecological niche space for diversification compared with coastal areas. To facilitate the evolutionary interpretation of geographic variation in diversification rate, we compared spatial diversification differences between montane and coastal habitats. First, Mantel tests indicated significant correlations between pairwise genetic and geographic distances for all montane lineages but did not reveal significant correlations in the coastal lineages *Talorchestia* and *Platorchestia* (table S14). Second, the bivariate and density plots demonstrated substantially higher genetic diversity in short distances for montane lineages, with multiple endemic species coexisting within a small mountain region (Fig. 2D and figs. S9 and S10). Third, the diversification analyses revealed that each montane lineage had a similar near-constant lineage accumulation trajectory and experienced higher initial but gradually decreased net speciation rates; this may be related to the occupation of new habitats by incipient species (fig. S11 and table S15). The spatial diversification rate, derived from the net diversification rates (λ-μ) from the binary-state speciation and extinction (BiSSE) (*11*) and λ_{BAMM} from Bayesian analysis of macroevolutionary mixtures (BAMM) (*12*) divided by their area size (λ-$\mu_{Montane}$ or $\lambda_{BAMM\text{-}Montane}$/166 grid cells, λ-$\mu_{Coastal}$ or $\lambda_{BAMM\text{-}Coastal}$/278 grid cells), showed faster spatial diversification rates in the montane lineages compared with coastal lineages (fig. S12 and table S16).

To test if the top-down cladogenesis and high endemism of montane lineages deviated from random expectation, we ran 1000 simulations of a phylogenetic parameter (the slope of the linear regression between the habitat elevation and divergence times) and geographic parameter (the slope of the linear regression between

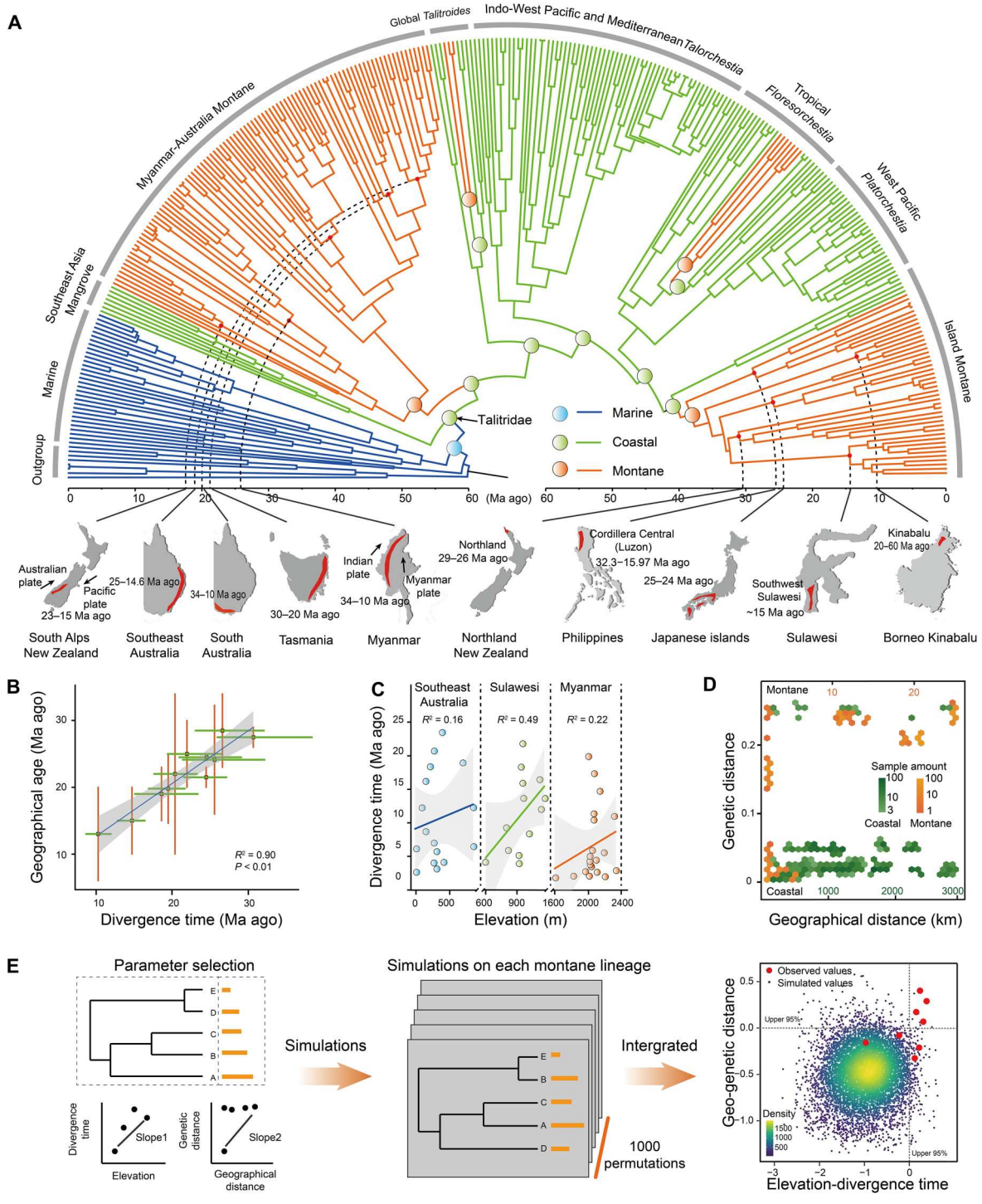

Fig. 2. Global marine-montane evolution of talitrids. (**A**) Dated phylogeny of talitrids. Clade colors refer to different habitat types. Circles on major nodes represent potential habitats inferred from ancestral reconstruction. The bottom maps represent geological distribution of montane lineages and geological ages of related transition events. (**B**) Linear regression and Spearman correlation between divergence times of major montane lineages (green horizontal bars contain 95% highest posterior density) and the corresponding geographical ages (orange vertical bars contain the upper and lower extremes of geological events). (**C**) Positive relationships between elevation and divergence time of internal nodes for montane lineages influenced by different geological events in Myanmar (plate collision), Sulawesi (island formation), and Southeast Australia (volcanic activity). (**D**) Bivariate plots of pairwise geographical and genetic distances show high genetic divergence within small geographic ranges for montane species (represented by *Myanmarorchestia* from Mount Victoria, Myanmar) and low genetic divergence for widely distributed coastal species (represented by the *Talorchestia* lineage from the Indo-West Pacific). (**E**) Distribution and diversification simulations for montane lineages. Parameters selected from phylogenetic and geographic perspectives. For the simulations, 1000 permutations were performed using random phylogenies and geological distributions. Eight montane lineages were integrated via simulated values that were normalized by the upper limits of the 95% confidence interval (CI) as follows: (slope − CI)/CI.

pairwise genetic and geographic distances) with random rearrangement of species coordinates and phylogenetic relationships (Fig. 2E). Simulation results showed that observed values from six of eight montane lineages fell near the upper limits of the 95% confidence interval (CI; figs. S13 and S14), indicating that patterns of top-down cladogenesis and montane endemism unlikely emerged at random.

Convergent gill innovation during montane transition

To test whether transitions to montane habitats were associated with a set of convergent morphological traits that contributed to high fitness, we assembled a morphological dataset (including body size and gill shape) from 569 individuals; this dataset included 49 marine, 112 coastal, and 158 montane species (Fig. 3A, fig. S15, and tables S17 to S19). Sensitivity analyses indicated that there was no significant difference in body length between species from different habitats (Fig. 3B). The principal components analysis (PCA) performed on Procrustes coordinates from 102 homologous landmarks generated a morphospace defined by two principal components. PC1 and PC2 provided the best approximation for the total gill shape variance and, together, accounted for more than half (59.9%) of the variation (Fig. 3C). Coastal species occupied the positive range of PC1, whereas montane species were distributed in the negative range, and significant differences existed between these two habitats ($P = 2.2 \times 10^{-16}$; Fig. 3C). According to the wireframe analyses, negative PC1 values reflected extra lobes and extra folds of the gill, whereas positive values indicated a streamlined and smooth margin (fig. S15C). The Procrustes analysis of variance (ANOVA) and phylogenetic ANOVA further indicate that the gills of marine, coastal, and montane species are significantly different from each other (table S20).

According to the analyses on phenotypic variance and habitat, a negative relationship was detected between gill shape (represented by PC1 value) and elevation, whereas a positive relationship was detected between gill size (represented by the centroid size, normalized by body length) and elevation (Fig. 3D), which indicated that montane species have more convoluted and enlarged gills than coastal species (fig. S16). Similar results were found between phenotypic variance and annual mean temperature, which is a climate variable that differentiates montane and coastal habitats (Fig. 3E, figs. S16 and S17, and tables S18 and S19). These results indicate the presence of convoluted and enlarged gills in colder montane habitats compared with relatively warmer coastal habitats.

The combined PCA plot of extant and ancestral species (fig. S15D) indicated that the gill shape of the MRCA of coastal and montane species had a positive PC1 value and was close to those of extant coastal species, whereas the gill shape of ancestral montane species had a negative PC1 value and was similar to those of the extant montane species.

The pairwise Procrustes distance analysis, weak phylogenetic signal ($K = 0.216$), and phylogenetic ANOVA among different lineages revealed that the morphological distances were not correlated with phylogenetic relationships (tables S20 and S21). The reconstructed phylomorphospace demonstrated evidence for convergent evolution because species from the same habitat occupied similar spaces (fig. S16E). The same instances of convergence became apparent in the tanglegram of the phenogram and phylogram, which showed clear clusters of coastal and montane clades, and converging morphological species from different clades occupied the same habitat (Fig. 3F and fig. S15E). Distance-based convergence measures also supported shape convergence in marine, coastal, and montane habitats (all $P < 0.001$; table S22).

Genetic changes of gills in montane habitats

To investigate genetic adaptation to montane life and convergent genomic change among distantly related montane clades (figs. S18 and S19), 1749 one-to-one single-copy orthologs were identified on the basis of genomes and transcriptomes from three montane, seven coastal, and two marine talitrid species. The phylogenomic tree inferred by linked orthologs was identical to the topology from the 851-tip full dataset and 397-tip species-level dataset (fig. S18A).

We obtained 274 positively selected genes (PSGs) across all three montane species using the branch-site likelihood ratio test in Phylogenetic Analysis by Maximum Likelihood (PAML) (table S23). Gene Ontology (GO) categories associated with PSGs across the montane species include "GO: 0006979, response to oxidative stress," "GO: 0005524, ATP binding," and "GO: 0055114, oxidation-reduction process" (details in fig. S18B).

In the gene expression analyses, a tissue-specific expression pattern was revealed by PCA and the hierarchal-clustering expression profile across all four species clustered by gills, brains, legs, and antennae (fig. S19, A and B). We identified 49 up-regulated differentially expressed genes in the gills relative to the other three tissues in montane species, compared with coastal species (fig. S19C). Details on the differentially expressed genes of montane species in each tissue can be found in fig. S19 (C to H).

Among the PSGs, 17 genes showed up-regulated expression in the gills of montane species (table S24). According to functional annotation, three of them participate in the process of energy metabolism (Fig. 3G); *AMY1* is involved in the first step in the digestion of dietary starch and glycogen (*13*). *OGDH* encodes proteins that catalyze the oxidative decarboxylation of α-ketoglutarate to reduce nicotinamide adenine dinucleotide (oxidized form) (NAD) to reduced form of nicotinamide adenine dinucleotide (NADH) during the tricarboxylic acid cycle (*14*), and *ACLY* encodes the primary enzyme responsible for cytosolic acetyl-CoA (coenzyme A) synthesis and participates in the lipogenesis pathway (*15*). Upon hypoxia/hypothermia retreatment of *M. aosen* (a species that lives at a low elevation and in a warm habitat), we found that *AMY1* was up-regulated in gills and legs regardless of treatment. The other two genes (*ACLY* and *OGDH*) only showed up-regulated expression in hypoxia treatment and had significantly increased expression in gills but not legs during hypoxia treatments (Fig. 3H).

As the evolutionary rate differences indicated 17 genes showed a significantly higher dN/dS ratio in montane species compared with coastal species (table S25). Three of them (*ENPP*, *PDHX*, and *HXK2*) are involved in the process of energy metabolism (Fig. 3G). *ENPP* plays a central role in purine signaling regulation by sequentially hydrolyzing adenosine triphosphate (ATP) to adenosine diphosphate to adenosine (*16*). *PDHX* encodes a component of the pyruvate dehydrogenase (PDH) complex that catalyzes the conversion of pyruvate to acetyl-CoA, thereby connecting glycolysis to the tricarboxylic acid cycle (*17*). *HXK2* encodes hexokinases, which catalyze the first essential step of glucose metabolism, the conversion of the substrate glucose into glucose-6-phosphate; this increases glycolysis rate (*18*).

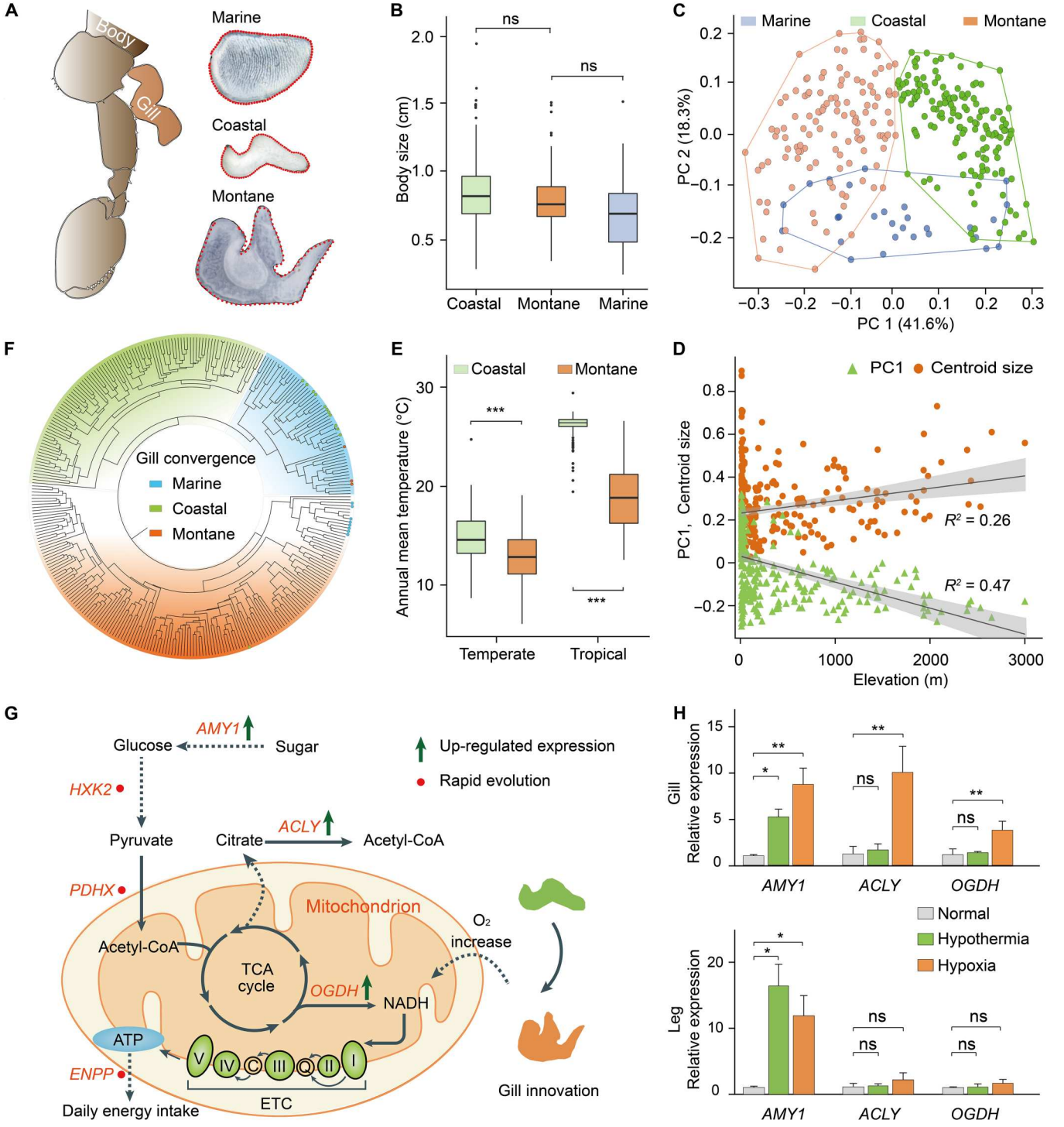

Fig. 3. Gill innovation, convergence among habitats, and related genetic changes of talitrids. (**A**) Gills of male gnathopod II from marine, coastal, and montane habitats. Homologous landmarks used for shape comparison are shown by red dots. (**B**) Comparison of body size (represented by body length) between species from different habitats. (**C**) PCA plot showing gill shape distribution. Numbers in brackets indicate the percentage of variance explained by each of the PC axes. (**D**) The negative relationship between elevation and PC1 (a proxy for gill shape) and positive relationship to centroid size (normalized by body size). (**E**) Comparison of annual mean temperature shows montane areas with colder temperatures than coastal habitats in both tropical and temperate regions. (**F**) UPGMA phenogram constructed from Procrustes distances showing convergent morphological traits of species in coastal and montane clades. Dots at the tips of the phenogram represent particular lineages with convergent morphology from different habitats. (**G**) Schematic displaying the PSGs that showed up-regulated expression or rapid evolution and are involved in the energy metabolism process. Gill innovation with extra lobes and extra folds increased oxygen supply during respiration. (**H**) Significantly increased expression of *AMY1*, *ACLY*, and *OGDH* in gills under hypothermia (10°C) or hypoxia (13.5% oxygen) treatment relative to normal conditions but nonsignificant expression changes of *ACLY* and *OGDH* in legs under hypoxia treatment. **P* < 0.05, ***P* < 0.01, and ****P* < 0.001 (Wilcoxon test). ns, not significant.

SMC3 function and knockout experiments

We identified 258 genes with common amino acid substitutions in any two of the three montane species. Higher semantic similarities existed between these convergent genes and PSGs of montane species based on functional enrichment analysis (fig. S18, B to D), which indicates that the genetic convergence among montane clades resulted from similar habitat pressures. Of these genes, 44 are considered adaptively convergent genes that have been subject to positive selection and undergone nonrandom convergent changes (table S26).

According to annotations of the adaptively convergent genes, we found that *SMC3* is involved in insect wing morphogenesis (*19*). Considering potential serial homologies between insect wings and crustacean gills (*20*, *21*), we hypothesized that the gene *SMC3* is related to gill morphology difference in talitrids. We first amplified the *SMC3* sequence from 45 montane, 50 coastal, and 9 marine talitrids and confirmed that an amino acid mutation at the sequence level only existed between habitats rather than clades (Fig. 4A). On the basis of these sequences, accelerated dN/dS values were detected on the nodes where habitat transition occurred compared with those without habitat transition, which indicated that this gene faced strong selective pressure during montane transitions (Fig. 4B). The three-dimensional structures of *SMC3* also demonstrated that the fixed mutation was predicted to be a ligand binding site and there were structural deviations between the three habitats (Fig. 4C and fig. S20). The gene expression analyses indicated that *SMC3* and *cut* were expressed at lower levels in gills of montane species than those of coastal species (Fig. 4D).

To investigate the phenotypic consequences of the *SMC3* mutation in crustaceans, knockout experiments using CRISPR-Cas9 were performed in *Daphnia magna*, which has a nearly worldwide distribution and short life cycle, and has been a model system in ecology and evolution of Crustacea for hundreds of years (*22*). We designed three single-guide RNAs (sgRNAs) and did not observe allele mutations for sgRNA-1 and sgRNA-2. However, for sgRNA-3, 25 mutants showed short, frameshift-inducing indels relative to wild-type sequences, and the majority showed that nonsense alleles caused frameshift mutations with a 17-bp insertion (Fig. 4E and fig. S21). In the mutant *Daphnia*, we observed four individuals with carapace defects, mainly irregular margins (Fig. 4F and fig. S21). Further quantitative PCR (qPCR) validation indicated that the mutants had significantly lower expression levels of *SMC3* and *cut* compared with wild-type in the carapace but did not have significant expression differences in the legs (Fig. 4, G and H).

DISCUSSION

Global marine-montane transition

Our informative genome-wide data provide a robust and well-resolved phylogeny and temporal framework to explore the evolutionary trajectory of talitrids (Fig. 1). Talitrids originated in the sea, with the first transition to coastal mangroves occurring approximately 56.16 Ma ago (Fig. 5). The expansion of mangrove forest since the early Eocene provided a suitable habitat and sufficient food sources for marine ancestral talitrids to colonize coastal habitats (*23*). Similar scenarios occurred in terrestrial gastropods and crabs, which colonized land from the sea via mangrove swamps (*24*). The coastal habitats experienced repeated tidal influences but allowed a gradual accumulation of terrestrial adaptations for the ancestral coastal talitrids, which inhabited cosmopolitan seashores and then independently shifted to mountains in different regions during the Oligocene to Miocene (Figs. 2A and 5).

Montane talitrids are distributed in mountains of continental margins or island arcs, where preexisting marine environments were uplifted by a series of geological changes. We summarized related geological marine-montane transition events of 11 mountains or islands at a global scale from the Oligocene to Miocene (table S10). The divergence times of three montane lineages corresponded well to the age of plate collision (table S10): the Myanmar lineage diverged approximately 26.6 Ma ago (when the Burma Plate collided with the northeast edge of the Indian plate) (*25*), the New Zealand Southern Alps lineage diverged approximately 18.6 Ma ago (the Southern Alps straddles a major tectonic boundary between the Pacific and Australian plates) (*7*), and the Sundaland lineage diverged approximately 24.46 Ma ago (when Australia moved north and was subducted under Sundaland) (*8*). The formation of five lineages was consistent with island formation processes (table S10): the Central Cordillera of the Philippines lineage formed ~25.56 Ma ago (as a result of the subduction of the South China Sea plate along the Manila Trench beginning in the Early Miocene) (*26*), the Southwest Sulawesi lineage ~14.72 Ma ago (the collision of different terranes in Sulawesi and subsequent fragmentation beginning ~15 Ma ago) (*27*), the Mount Kinabalu, Borneo lineage formed ~10.28 Ma ago (orogeny in the Kinabalu area began ~20 Ma ago after the collision of North Borneo with the South China continental crust) (*28*), the Northland New Zealand lineage ~30.65 Ma ago (Northland ophiolite terrane of northern New Zealand indicates that it formed in a suprasubduction zone setting) (*29*), and the Japanese islands lineage ~24.52 Ma ago (which coincides with when continental rifting made the Japan Sea push the Eurasian fringe out to form a series of offshore islands) (*30*). Two Australian lineages and the Tasmanian lineage originated later than the continent formation, but their divergence is concordant with some younger Oligocene volcanic orogenic events on the continental margins (*31*, *32*).

The divergence times of montane lineages correspond well with geological events that lead to marine-montane transitions (table S10), which indicates that geological events played a vital role in the biological transition of marine organisms moving into land areas (Figs. 2B and 5). Fossil deposits of marine gastropods, corals, and calcareous algae in montane areas of Southeast Asian islands and Australia also provided insight into geological marine-montane transitions (table S10).

For lineages that colonized montane areas, early divergent nodes were associated with higher ancestral elevation, and later divergent nodes were associated with lower ancestral elevation; this demonstrated a top-down cladogenesis pattern with a positive relationship between habitat elevation and divergence times (Fig. 2C). The underlying scenarios could be that early ancestors and their land substrates were elevated to montane top by orogenetic events, subsequently diversified downslope. This type of elevational diversification pattern can be found in lineages influenced by different geological marine-montane transition events, such as in Myanmar (plate collision), Sulawesi (island formation), and Southeast Australia (volcanic activity). With marine-montane transitions, mountains provide ecological opportunities for talitrid diversification and result in higher initial net speciation rates and occupation of new habitats by incipient species (figs. S11 and S12 and tables S15 and S16). The steepness of mountains and the ecological gradient

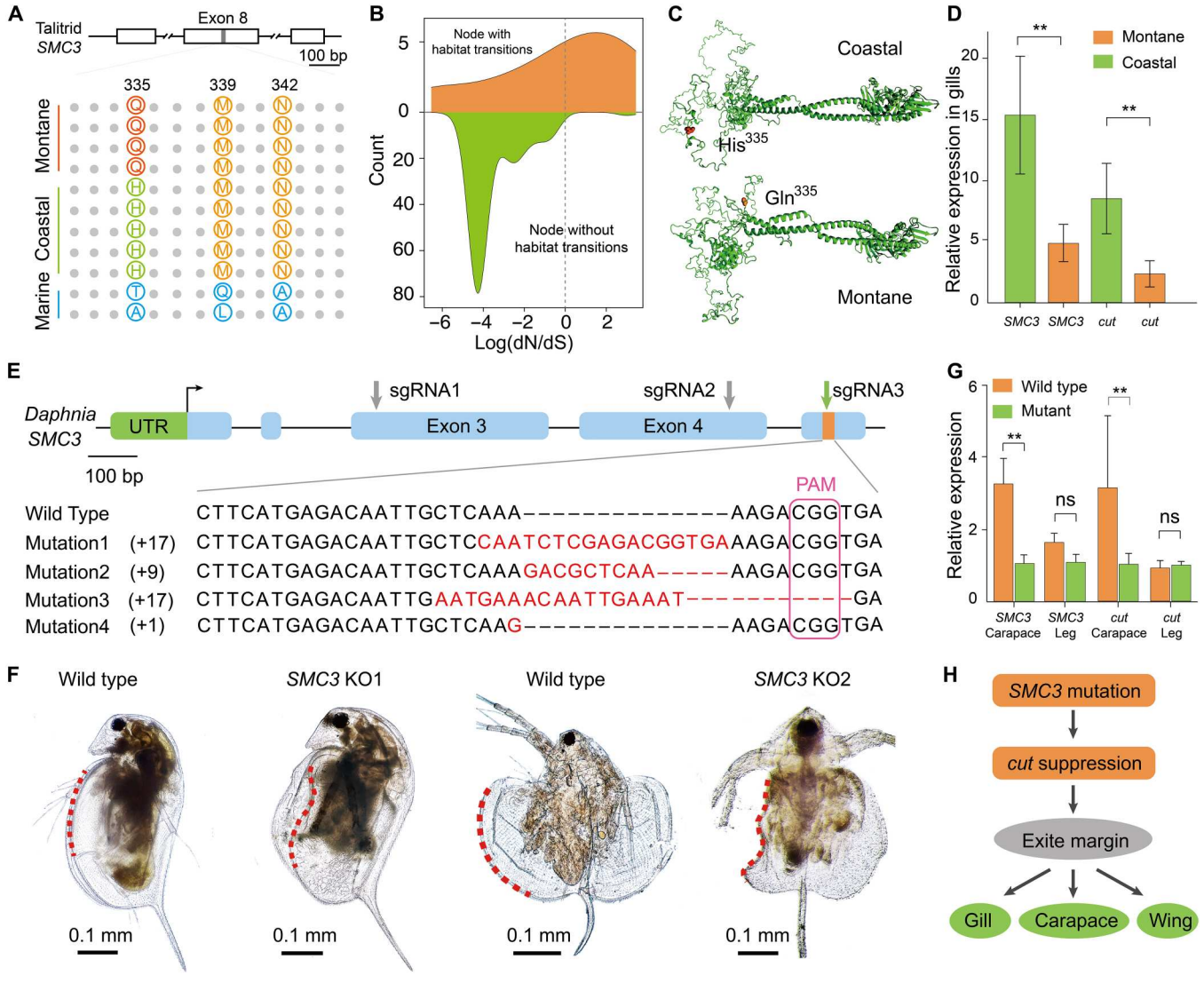

Fig. 4. *SMC3* regulates the exite margin development in crustaceans. (**A**) Comparison of amino acid substitutions in the *SMC3* protein sequence. Polymorphic and fixed substitutions are indicated with colored circles that correspond to different habitats. (**B**) Accelerated dN/dS ratio on the nodes where habitat transition occurred compared with nodes without habitat transition. (**C**) Three-dimensional structure models of *SMC3* in coastal and montane talitrids inferred by homologous modeling. The fixed substitution site was predicted to be a ligand-binding site and is depicted in red. (**D**) Significantly decreased expression of *SMC3* and *cut* in gills of montane species relative to coastal species. (**E**) Mutations generated in the CRISPR-Cas9 *SMC3* knockouts. The sgRNA target sites used for mutagenesis are indicated by the green arrow. Gray arrows represent the designed sgRNA; however, a stable genotype with allele mutation was not obtained. The short, frameshift-inducing indels relative to the wild-type sequences were identified by PCR, molecular cloning, and Sanger sequencing. *SMC3* sgRNA targets are highlighted in red, and PAM sequences are highlighted in pink. (**F**) Notable carapace margin defects were found in homozygous *SMC3*-knockout *Daphnia*. (**G**) Expression patterns of *SMC3* and *cut* in *D. magna* revealed significantly decreased expression of these genes in the carapace of *Daphnia* mutants relative to the control and without significant expression shifts in the leg after *SMC3* knockout. (**H**) Potential regulatory pathway of *SMC3* from suppressing *cut* that affects the exite margin development, which leads to phenotypic changes of the talitrid gill, *Daphnia* carapace, and insect wing. $^{**}P < 0.01$ (Wilcoxon test).

increase habitat diversity per geographic unit, which leads to numerous endemic species that aggregate and diversify in the geographical vicinity. This results in higher genetic diversity and spatial diversification rates (Figs. 2D and 5 and figs. S9 to S12), forming highly diverse centers of montane endemism and underestimated biodiversity (*33*).

The permutation models revealed that the observed patterns of montane lineages (six of eight) were significantly different from the simulated random dispersal pattern, and most of these lineages (excluding the South Australia and Northland New Zealand lineages) were distributed near the upper limits of the 95% CI. It is possible that the South Australia and Northland New Zealand lineages were outside the 95% CI because of incomplete sampling and overlapping ranges. For example, samples from Northland New Zealand are limited from 50 to 130 m in altitude. These results indicated that the current montane species distribution reflected by our sampling and phylogeny did not result from random dispersal, which further supports the occurrence of top-down cladogenesis.

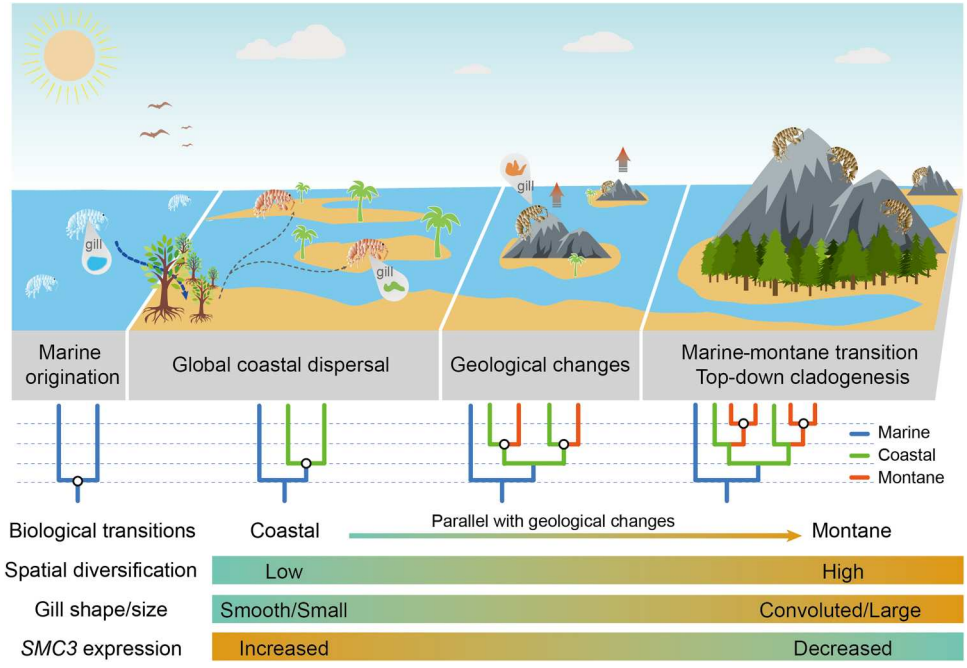

Fig. 5. Temporal process and corresponding cladogenesis of marine-montane transition in talitrids. The biological transitions from coastal to montane habitats occurred in parallel with geological changes and were coupled with diversification, gill, and genetic changes.

Convergent gill innovation and related genetic changes

During marine-montane transitions, respiration should be the first process to change, and marine ancestors often evolved morphological innovations for air breathing (*1*, *2*). PCA and Procrustes ANOVA demonstrated a significant difference in gill morphology among species from different habitats; montane species have larger gills with a more convoluted shape than coastal species (Figs. 3C and 5, fig. S15C, and table S20). The correlation analyses indicated that this variance was significantly correlated with habitat variables, e.g., elevation and annual mean temperature (Fig. 3, D and E, and fig. S16, C and D). Ancestral reconstruction of gill shape showed that the montane gill shape was coupled with montane colonization, which indicated that gills changed during the montane transition (fig. S15D).

Phylogenetic ANOVA, the reconstructed phylomorphospace, distance-based convergence measures, and phenogram all demonstrated the convergent evolution of gills on global mountains (Fig. 3F, fig. S15E, and tables S20 to S22). The convergent enlarged gills with extra lobes and extra folds increase oxygen supply during respiration. In addition, montane species appear to have a thickened gill cuticle to prevent desiccation (*9*). These variations of gill shape may be related to maintenance of daily energy intake in colder environments for montane talitrids and may have been beneficial for global habitat transitions.

The genomic and transcriptomic comparisons and hypoxia/hypothermia acclimation experiments provide further understanding of energy metabolism when an enlarged gill is supplied with sufficient oxygen. Selective pressure and differentially expressed gene analysis revealed that three PSGs showed up-regulated expression in the gills (*AMY1*, *OGDH*, and *ACLY*), and these genes are involved in the process of energy metabolism (Fig. 3G and tables S23 and S24). Hypoxia and hypothermia treatment of coastal species validated the up-regulated expression of these genes in gills when responding to montane environments, which indicates that these genes participate in hypoxia response and increase expression in the gills (Fig. 3H). In addition, three rapidly evolving genes (*ENPP*, *PDHX*, and *HXK2*) are also involved in this process, which demonstrates accelerated selection of this process across montane species (table S25). Similar energy metabolism has been observed in terrestrial crabs and seems to be a useful genetic response to terrestrial living in crustaceans (*34*).

The role of *SMC3* in gill innovation

Convergent phenotypes could be driven by convergent genetic changes. Among the adaptively convergent genes (table S26), *SMC3* shows a fixed amino acid mutation at the sequence level between habitats and an accelerated dN/dS ratio on the nodes with habitat transitions. Protein structural prediction verified three-dimensional structural variations between habitats. Furthermore, significantly lower expression levels of *SMC3* and *cut* were detected in gills of montane species. Collectively, these results reveal that *SMC3* is a crucial gene involved in habitat transitions and gill morphogenesis.

SMC3 encodes cohesion protein binding to the regulatory region of *cut*, which encodes a homeobox transcription factor; this suppresses *cut* expression in the developing wing margin (*19*). *SMC3* knockdown in fruit flies results in the loss of the wing margin (*19*). The insect wing has been proposed to be potentially serial homology to crustacean gills (*20*, *21*). Bruce and Patel (*35*, *36*) documented that outgrowths, such as insect wings, crustacean gills, and the *Daphnia* carapace, originated from ancestral crustacean exites (lateral leg lobe), which are ready to be de-repressed/activated under the right circumstances and then molded into new shapes and functions by evolution. On the basis of the hypothesis that

gills, wings, and carapace derived from exites, we propose that this gene has a potential regulatory role in the talitrid gills.

Our knockout experiments using CRISPR-Cas9 in *D. magna* showed that the mutant *Daphnia* had notable carapace defects, mainly from irregularly shaped margins (Fig. 4, E and F). Further qPCR validation demonstrated significantly decreased expression of *SMC3* and *cut* in *Daphnia* mutants (Fig. 4G), with similar scenarios for montane talitrids. Therefore, we propose that *SMC3* has a possible function in regulating morphological changes that influence the development of the appendage exite margin, including the talitrid gill, *Daphnia* carapace, and insect wings. The regulatory pathway of *SMC3* is likely related to down-regulated expression of *cut* and further involved in talitrid gill development (Fig. 4H). Future research on talitrids using CRISPR-Cas9 will provide a comprehensive explanation of *SMC3* function in gill development.

In summary, we report Oligocene to Miocene global marine-montane transition of crustaceans and demonstrate geological changes related to the biological transition that triggered top-down cladogenesis (Fig. 5). Convergent gill evolution and the regulatory mechanism of *SMC3* demonstrate that respiratory innovation may have facilitated the marine-montane transition of talitrids. The key trait innovations and related genetic mechanisms account for the evolutionary success of talitrids′ global diversification and elucidate the formation of prosperous terrestrial biodiversity.

MATERIALS AND METHODS

Taxon sampling and DNA sequencing

In this study, we sampled 851 talitrids worldwide and covered most of the distribution of the family (table S1). Almost 86.3% of the samples (734 of 851) were collected from 592 sites from our group and several collaborators over three decades, another 13.7% samples (117 of 851) were downloaded from National Center for Biotechnology Information (NCBI). We defined habitats of talitrids as marine, coastal, and montane. Marine species are those commonly found swimming in seawater. Coastal species depend on algal litter and beach debris and can swim in water, although they are usually found jumping on the seashore. Montane species are inhabitants of forest leaf litter and accustomed to a fully terrestrial existence, and they exhibit a unique jumping behavior. Twenty-seven samples were selected as outgroups mainly based on the data available from the work of Hou and Sket (*37*). Detailed sampling information is given in table S1.

All procedures performed on animals were approved by the Animal Care Committee of the Institute of Zoology, Chinese Academy of Sciences. DNA was extracted from legs of specimens using the TIANamp Genomic DNA kit (TIANGEN Inc., Beijing, China). Three mitochondrial fragments of cytochrome c oxidase subunit I (*COI*), *COII* and 16*S* ribosomal RNA, and four nuclear regions of 18*S*, 28*S* ribosomal RNA, histone 3 (*H3*) and sodium-potassium adenosine triphosphatase a-subunit (*NaK*) were amplified. Details on primers and PCR conditions were provided in table S2.

Genome and transcriptome sequencing, assembly, and annotation

Specimens of *M. aosen* were collected from Olympic Park, Beijing, China. Genomic DNA for short and long read sequencing was isolated from 12 adult talitrids from a same breeding lineage, using the QIAGEN Blood & Cell Culture DNA Kit (QIAGEN, Hilden, Germany). A total of 239.88-Gb clean reads produced on the DNBseq platform were used for genome survey and assembly correction after removal of low-quality reads and adapters. A total of 118.46 Gb of HiFi reads were generated for genomic contig assembly using hifiasm v0.15.2 (*38*) with default parameters. The 266-Gb reads from the Hi-C library sequencing were used for chromosome construction after primarily filtered and mapped to the genome.

Another five species (*Vallorchestia* sp., *Cochinorchestia* sp., *Floresorchestia mkomani*, *Talorchestia martensii*, and *Platorchestia pacifica*) were selected for short read genome sequencing and draft genome assembly using SPAdes v3.13.0 (*39*) and K-mer values of 21, 41, 61, 81, 101, and 121 were selected, based on around 300-Gb clean reads produced from DNBseq platform.

Genome quality and completeness assessment was conducted with BUSCO v4.1.1 (*40*) using the arthropoda_odb10 and generated genome statistics using the assembly-stats v1.0.1 (https://github.com/rjchallis/assembly-stats).

For gene annotation of genome, we sequenced the transcriptomes of whole body and different tissues from *M. aosen*, including the antennae, brains, gills, and legs. The total RNA of the whole body from one individual, and the tissue RNA from a mixture of tissues from 10 to 15 individuals, were extracted using the RNAsimple Total RNA kit (TIANGEN, Beijing, China). Sequencing was performed on the NovaSeq 6000 platform, and each transcriptome got at least 6-Gb clean data after removing reads containing adapter, ploy-N, and low-quality reads. For genetic analysis, four tissues same as above from three species (*Floresorchestia xueli*, *Myanmarorchestia nunomurai*, and *P. pacifica*) were sequenced. Clean reads were quality checked and trimmed using the FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) and FASTXToolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). We also sequenced two full-length transcriptomes for talitrid species of *M. nunomurai* and *F. xueli*, based on the PacBio Sequel platform. A total of 33.62-Gb clean data were obtained for *F. xueli*, and 33.37 Gb were obtained for *M. nunomurai*.

Three strategies were applied for gene structure prediction across the genome: homology-based annotation, de novo prediction, and RNA sequencing (RNA-seq)–based annotation. For homology-based annotation, the published genomes of three amphipods—*Hyalella azteca* (GCA_000764305.3), *Trinorchestia longiramus* (GCA_006783055.1), and *Parhyale hawaiensis* (GCA_001587735.2)—were used by GeMoMa (*41*). For transcriptome-based prediction, the RNA-seq reads from different tissues were aligned to the genome to identify the exonic regions and splice positions to optimize the annotation. We identified and classified repetitive elements de novo and generated a library of consensus repeat sequences for the genome using RepeatModeler (www.repeatmasker.org/RepeatModeler/). We then annotated and masked repeats in the assembly with RepeatMasker v4.1.2 (www.repeatmasker.org/) using the custom repeat library generated in the previous step. Besides, Tandem Repeats Finder (TRF) v4.09 (*42*) was used to analyze tandem repeats. Genes were functionally annotated against NR (www.ncbi.nlm.nih.gov), Swiss-Prot (www.uniprot.org/), GO (http://geneontology.org/).

Phylogenomic marker design

To establish a solid phylogenetic framework, we designed a series of primers to produce numerous genomic markers suitable for

Amphipoda based on the generated chromosome-level genome, five draft genomes, and publicly available genome sequences of *H. azteca* and *P. hawaiensis*. Transcriptomes of two amphipod species (*Gammarus electrus* and *Jesogammarus hebeiensis*) were added to expand the scope of the application of primers in the order Amphipoda.

First, single-copy and long exons were selected on the basis of the genome of *M. aosen*. Then, exons from *H. azteca*, *P. hawaiensis*, and five draft genomes were used for reciprocal best-hit (RBH) (*43*) with the selected exons, with identity >70% and coverage >50%. If the selected exons BLAST hit more than three species, then they were treated as an orthologous sequence for primer design. We found 268 orthologous sequences that were used to design primers as candidate genes. All the candidates were aligned and trimmed on both ends according to the sequences of *M. aosen*, *H. azteca*, and *P. hawaiensis*. The aligned candidate genes were displayed at the amino acid level, and conserved blocks (>7AA) were selected to design nested PCR primers to improve amplification success.

Last, a total of 192 pairs of primers for 96 candidate genes were designed, and all of these were amplified using nested PCR (table S8). The first-round PCR cycling program was an initial denaturation step of 94°C for 4 min, 35 cycles of 94°C for 45 s, 45°C for 40 s, 72°C for 2 min, and a final extension step of 72°C for 10 min. The second-round PCR cycling program was an initial denaturation step of 94°C for 4 min, 30 cycles of 94°C for 45 s, 50°C for 40 s, 72°C for 1.5 min, and a final extension step of 72°C for 10 min. All the obtained amplification products from the same taxon were pooled together; then randomly sheared to small fragments (200 to 500 bp) by dsDNA Fragmentase (NEB), then blunt-end-repair, and A-tailing; and attached to a species-specific barcode linker. We mixed 24 indexed amplification product pools together and sequenced them on an Illumina HiSeq 2500 platform. About 3 Gb of Illumina HiSeq paired-end 150-bp data were obtained. The raw sequence was sorted by barcode into different taxa. Each species' data were then assembled into full-length sequences independently. The assembled sequences were examined by BLASTn searches against the 96 reference sequences from the primer design. Last, all of the refined sequences were checked again to see whether they could be correctly translated.

Biomarker performance and phylogenetic reconstruction

All 96 NGS nuclear markers and seven Sanger sequencing markers were aligned by the MAFFT v7.455 with the G-INS-i model (*44*). To test the performance of biomarkers, we calculated variable sites for 96 NGS markers and compared them to the Sanger sequencing markers.

We performed GMYC analysis to define the MOTUs by using R package splits (*10*) based on the ultrametric tree generated with mitochondrial genes for all samples in BEAST v2.5.2 (*45*). Mitochondrial sequences for 12 samples are unavailable, but their morphology characters are different from existing MOTUs; therefore, we treated them as independent MOTUs. We chose one sample from each MOTU (the ones with less missing genetic data) and composed a 397-tip dataset at the species level (table S1).

For phylogeny reconstruction, ML analyses were first carried out using RAxML v8.0 (*46*). Branch support was evaluated with 1000 rapid bootstrap replicates. A thorough optimization of the best-scoring ML tree and the GTRGAMMA model of rate heterogeneity were applied. We also conducted ML analyses and used the embedded ultrafast bootstrap approach to estimate nodal support values using IQ-TREE v1.6.12 (*47*). Bayesian analyses were conducted using ExaBayes v1.4.1 (*48*). Two Markov chain Monte Carlo (MCMC) runs were performed with one cold chain and three heated chains for 50 million generations, sampling every 1000 generations. The average standard deviation of split frequencies and potential scale reduction factor were <0.01 and close to 1 across the two runs, respectively. The effective sample sizes were checked, and all parameters were >200 after the first 20% of generations were discarded (*45*). The species tree analysis without gene concatenation was performed with ASTRAL v4.7.6 (*49*) under the coalescent model based on the species-level dataset. The species tree analysis was conducted using the best ML trees and corresponding bootstrapping trees for each of 103 genes inferred by RAxML under the multilocus bootstrapping option with 1000 replicates ($-r$ = 1000).

The pairwise distances between ML tree topologies generated from the single-gene, and the concatenated topologies were calculated on the basis of Robinson-Foulds distance (https://scriptomika.wordpress.com/2014/01/27/59/). The tree-to-tree distances were then visualized using multidimensional scaling in the R package reshape2 (*50*).

Divergence time estimation

For molecular dating, we used two fossils and five geological events as calibration points: (i) The MRCA of *Tethorchestia* and *Mexorchestia* was set at ~16 Ma ago (lognormal, offset = 16, M = 2, S = 1) based on Dominican amber fossils of the talitrid *Tethorchestia* (*51*). (ii) We constrained the MRCA of two endemic *Palmorchestia* species at ~2 Ma ago (normal, mean = 2, sigma = 0.1) assuming that the clade originated simultaneously to or later than the formation of La Palma (Canary Islands) (*52*). (iii) The Eocene Baltic amber record of *Niphargus*. We retrieved *Niphargus* sequences from GenBank (table S1) and constrained the MRCA as ~45 Ma ago (lognormal, offset = 45, M = 2, S = 1) (*53*). (iv) The separation of Arctic marine and freshwater *Gammaracanthus* was connected with Plio-Pleistocene climatic cycles at ~6 Ma ago (normal, mean = 6, sigma = 1) (*37*). (v) The MRCA of the Baikal amphipod clade, for which we assigned ~27 Ma ago (normal, mean = 27, sigma = 2) based on the inception of rifting in the Baikal Basin triggered the formation of Lake Baikal around 35 to 27 Ma ago (*54*). (vi) The MRCA of *Sarothrogammarus* and *Rhipidogammarus* originated at ~37 Ma ago (normal, mean = 38, sigma = 2) corresponding to the regression of the Tethys from east Pamir in the late Eocene (*55*). (vii) The onset of the *Gammarus* radiation in Lake Ohrid was constrained to ~1.36 Ma ago (normal, mean = 1.36, sigma = 0.1) based on the formation of the ancient lake ecosystem (*56*). Calibration nodes four to seven from outgroup have been widely used in dating amphipod phylogenies (*37*). We constrained the maximum age of the root to 358 Ma ago to represent the origin of Amphipoda based on the earliest known Eumalacostraca fossil from the Late Devonian, *Palaeopalaemon newberryi* Whitfield, 1880 (*57*).

We first used the seven Sanger sequencing markers to estimate the divergence times in BEAST2 with an uncorrelated lognormal relaxed molecular clock model with a Yule process for the speciation model and GTR + I + G for the substitution model based on 851-tip full dataset and 397-tip species-level dataset obtained from IQ-TREE. Next, we used penalized likelihood method as

implemented in treePL v1.0 (*58*) and conducted two analyses independently based on topology from full data and species-level data. Last, we estimated divergence times using Bayesian relaxed clock approaches implemented in MCMCTREE from the PAML package v4.9j (*59*, *60*) with the uncorrelated rate model (clock = 2). The phylogram of the species-level data was used as the reference topology.

To test the sensitivity of our time calibration to particular calibration points, we used a jackknife analysis based on the BEAST2 calibration. We successively deleted each of the individual calibration points and repeated the calibration analyses with the same setting parameters. The correlation of median age estimates at major nodes between these analyses, and the full calibration was assessed with Spearman's rank correlation analysis.

Inferences of ancestral ranges and ancestral habitats

To infer the ancestral ranges of talitrids, we used the DEC, DEC + j, DIVALIKE, DIVALIKE+j, BAYAREALIKE, and BAYAREALIKE+j models implemented in the R package BioGeoBEARS (*61*) based on the calibrated species-level tree from treePL, allowing dispersal without restriction. The Akaike information criterion and likelihood values were evaluated to test the fit of models. The distributions were categorized into seven biogeographic ecoregions: Australia region (including Australia, New Zealand, and Fiji), Indian Ocean (including Indian Ocean coast area), Mediterranean region (including Europe, Mediterranean Sea coast, and East Atlantic region), Southeast Asia (including Southeast Asia islands and coast area), Myanmar (including Myanmar inland region and southwestern China), Northwest Pacific region (including Japanese islands and Russian Far East), and eastern China.

We reconstructed the ancestral habitat with PastML (*62*) on the basis of the calibrated species-level tree using the marginal posterior probabilities approximation method with an F81-like model. Marine, coastal, and montane states were coded as discontinuous habitat traits. We also applied the "ace" function in the R package ape (*63*) to evaluate ancestral states across the phylogeny by generating 1000 stochastic character maps. The posterior probabilities were summarized at nodes to infer each node's best-supported ancestral state.

Diversification dynamics associated with habitat shift

On the basis of the calibrated species-level tree from treePL, we inferred the ancestral elevation at all internal nodes of each montane lineage using the "fastAnc" function in the R package phytools (*64*). The inferred ancestral elevation of the root was compared with the medium and average elevation of corresponding tips. Then, we independently extracted divergence times of internal nodes to infer the linear regression relationship between elevation and divergence time.

To assess the biodiversity of montane species, the Mantel tests were conducted to obtain a correlation between genetic distance and geographical distance using 1000 permutations to assess significance in the R package vegan (*65*). The pairwise genetic distances between the pairs of tips based on the branch lengths from the IQ-TREE analysis were calculated by "cophenetic.phylo" from ape. The pairwise geographical distance was calculated in the R package geosphere based on the sampling coordinates (*66*). The relationship between pairwise geographical and genetic distances was visualized by the bivariate plot and density plot.

We applied a series of methods to detect the diversification modes within lineages or between habitats. First, the lineage-through-time (LTT) plots for each montane lineage were constructed to visualize the diversification rate over time. The variation of the diversification rate shown on the LTT plots was calculated by the "bd.shifts.optim" function in R package TreePar (*67*). We used seven BiSSE models (table S16) implemented in the R package "diversitree" to model speciation (λ), extinction (μ), and transition (q) rates of coastal versus montane talitrid lineages (*11*, *68*). BiSSE enables correcting for incomplete sampling by indicating a sampling fraction (total, 58.05%; coast, 48.7%; mountain, 72.7%). This fraction was based on the number of our sampled species divided by already known species until October 2021 from the World Register of Marine Species (www.marinespecies.org). Last, we estimated the diversification dynamics of each clade and the speciation rates (λ_{BAMM}) for each tip using the BAMM v2.5.0 (*12*).

To better predict the diversification difference between coastal and montane habitats, we used an index of spatial diversification rate, defined as diversification rate weighted by range size (*69*). We first quantified distribution range area using 0.1° × 0.1° grid cells across the species distribution. We calculated the spatial diversification rate by dividing the net diversification rates (λ-μ) from BiSSE and λ_{BAMM} from BAMM by the range size (λ-μ, λ_{BAMM}/grid cells of range size) to detect the spatial diversity changes after the marine-montane transition.

Distribution and diversification pattern simulations

To determine whether the relationship between phylogeny and geographic distribution (including elevation and pairwise geographic distance) deviated from a random expectation, we simulated the distribution and diversification patterns in R using a bootstrap approach for the eight montane lineages described above independently. Two parameters were calculated and compared to our observed phylogenetic results: the slope of the linear regression between the distribution elevation and divergence times and the slope of the linear regression between pairwise genetic and geographic distances.

In our models, a random evolutionary history was generated (number of taxa and root age were fixed but phylogenetic relationships were random), and each species' geographic distribution (location and elevation) was randomly rearranged; one thousand permutations were performed for each montane lineage. Observed values from our phylogeny distributed within the limits of the 95% CI of simulations indicated that their distribution and diversification patterns were caused by random dispersal. Observed values greater than the upper 95% CI from an elevation-divergence time perspective revealed a positive relationship between elevation and divergence time, which indicated top-down cladogenesis. Observed values from a geographic-genetic distance perspective that were greater than the upper 95% CI indicated significant isolation by distance.

Climatic difference between montane and coastal habitat

To explore the climatic difference between coastal and montane species in statistics, we did a phylogenetic PCA using "phyl.pca" function in phytools based on the 19 bioclimatic variables and dated species-level phylogenetic tree. For species with multiple specimens, the mean for each of the 19 variables was calculated and used. To decrease the influence of latitude difference on the

temperature of sampling location, we divided tropical (latitude < 23.5° north or south) and temperate (latitude >23.5°north or south) ecoregions and analyzed them separately within each ecoregion. After removing those that were highly correlated (i.e., correlation > 0.9), ten variables were selected (table S18). All ten variables were used first, and then, temperature variables (bio1, bio2, bio3, and bio7) and precipitation variables (bio12, bio13, bio14, bio15, bio18, and bio19) were used separately for further analysis. The variables that can differentiate montane and coastal species significantly were chosen for future analysis.

Gill morphology analysis

We first dissected and photographed the gill of male gnathopod II from 569 individuals. Then, we measured the body length of specimens with a Vernier caliper to show body size differences between different habitats. Geometric morphometric analyses of gill variation were conducted on the basis of 102 landmarks (Fig. 3A) that were digitized with tpsDig v2.05 (*70*). The first two landmarks were marked as the two sides of the joint site with body and the other 100 landmarks resampled by equal length according to a curve drawn along the edge of the gill from left to right counterclockwise. All geometric morphometric analyses were carried out in the R package geomorph v3.0.5 (*71*) unless stated otherwise. All measured gills were scaled, translated, and rotated against the consensus configuration using generalized Procrustes analysis (GPA). Species with multiple specimens in the dataset were represented by their mean Procrustes shape and centroid size values.

To describe the patterns of shape variation, we used covariance matrices of shape data after GPA to calculate the principal axes of shape variation with PCA. PC1 and PC2 morphospace plots by wireframes provided visualization that facilitated identification of patterns of clustering and related gill shape changes. PC1 explained 41.6% of the gill variation and significant differences between habitats along PC1 values, so we used PC1 as a proxy for gill shape. In our study, gill size was represented by the centroid size, which was estimated as the square root of the sum of squared distances of each landmark from their centroid and normalized by body length. For statistical analyses of phenotypic variation, we compared the PC1 values and centroid size between separate habitats using Wilcoxon rank sum test. Procrustes ANOVA with 1000 random permutations of the residuals among habitats and clades was applied for significance testing. Furthermore, we estimated morphological disparity and performed pairwise comparisons among marine, coastal, and montane habitats and phylogenetic clades using "morphol.disparity" function. We estimated Blomberg's *K* (an estimator that assesses the strength of the phylogenetic signal in quantitative variable) using the "physignal" function to investigate the influence of phylogeny on gill shape variation (*72*). The linear regressions and correlation analyses were conducted between two habitat variables (elevation and annual mean temperature) and two phenotypic representatives (PC1 values and centroid size).

We reconstructed ancestral gill shape of all nodes using all landmark coordinates from the modules of the Rhetenor package in Mesquite v3.5 (*73*). All the calculated nodes' gill shapes were exported and integrated with landmark data of the tips as a combined dataset to infer gill difference among extant and ancestral talitrids. PCA was used to visualize morphospace occupation of ancestral gill shape.

A series of approaches were used to assess the scope and strength of convergence, and we hypothesized that species from the same habitat group would "converge" toward similar shapes. First, for an initial visual inspection, we produced a phylomorphospace plot in PC1-PC2 shape space. Second, to quantify convergent evolution in the different habitats, the distance-based convergence measures C1 to C4 were computed using the R package convevol (*74*). Significance was assessed in the same package using 1000 simulations along the phylogeny according to a Brownian Motion model. Third, the unweighted pair group method with arithmetic mean (UPGMA) phenogram was constructed on the basis of Procrustes distances from species mean configurations and computed as a tanglegram to compare the molecular phylogeny and phenogram using the "cophylo" function in phytools.

Genetic adaptation analysis

Nine genomes [including two marine species *H. azteca* and *P. hawaiensis* and seven coastal species *M. aosen*, *T. longiramus* , *P. pacifica*, *F. mkomani*, *T. martensii*, *Vallorchestia* sp., and *Cochinorchestia* sp.], two full-length transcriptomes (*F. xueli* and *M. nunomurai*), and one transcriptome (*Curiotalitrus* sp.) were used for detecting the genetic basis of montane adaptation. For identifying orthologous genes, the RBH method was used with an *E* value cutoff of 1×10^{-5} and a minimum percentage identity of 70% based on protein-coding DNA sequences (CDS). *M. aosen* was used as an anchor species.

To detect coding sequence changes involved in montane adaptation, we identified PSGs using the branch-site model and likelihood ratio test based on the well-aligned orthologs. Three montane species (*F. xueli*, *M. nunomurai*, and *Curiotalitrus* sp.) were set as foreground branches in PAML. The null model (model = 2, NSsites = 2, fix_omega = 1) and the alternative model (model = 2, NSsites = 2, fix_omega = 0) were used, and *P* values were calculated using chi-square distribution with one degree of freedom. A false discovery rate (FDR) cutoff of 0.05 was used in multiple test corrections.

For comparing evolutionary rate difference of PSGs between separate habitats, we used "Free-ratio model" on the branches in PAML to estimate the ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site (dN/dS) of each branch for all orthologs (method = 1, model = 1, verbose = 1). The difference of the dN/dS ratio between montane and coastal species was compared using the Wilcoxon rank sum test. Genes with a higher dN/dS value for the montane species than the coastal species were considered as evolving with a significantly faster rate.

For identifying differentially expressed genes across coastal and montane species during marine-montane transitions, the RNA-seq reads of four tissues (antennae, brains, gills, and legs) from two coastal species, *P. pacifica* and *M. aosen*, and two montane species, *F. xueli* and *M. nunomurai*, were mapped to their own trimmed orthologs. For each species, the expression level (fragments per kilobase million) of orthologs was calculated with RSEM v1.3.3 (*75*) and log-transformed ($N + 1$), following the quantile-normalization. Hierarchal-clustering expression profile analysis was conducted on the basis of Spearman's correlation coefficients of gene expression between pairs of samples. Differential gene expression between tissues and species was compared using "DESeqDataSetFromMatrix" and "DESeq" function in R package DESeq2 (*76*).

Up-regulated differentially expressed genes of gills from montane species with fold change >2 and 5% FDR were filtered for further analysis.

We examined convergent amino acid substitutions between pairs of three montane species by using JTT-f_{gene} amino acid substitution models according to branch model in the CODEML module of the PAML. A Poisson test was used to detect differences between the observed and expected number of convergent sites in each ortholog (*77*). Genes in which convergent amino acid changes in montane species significantly exceed expectations ($P < 0.05$) were considered convergent candidates.

The functional enrichment analyses of the PSGs and convergent genes were performed by KOBAS 3.0 (*78*) using Fisher′s exact test ($P < 0.05$) with *Drosophila melanogaster* as a reference. To evaluate the overall similarity of the functional enrichment patterns between the convergent genes and PSGs, we calculated pairwise semantic similarity (SS) values of the GO terms and examined SS values relative to random samples with the R package GoSemSim based on the BMA algorithms (*79*). To determine whether the observed SS values notable deviated from a random expectation, we performed two different permutations. First, we randomly sampled the same number of genes (274 PSGs and 258 convergent genes) from the pool of all 1749 genes. We repeated this sampling 1000 times and then performed analogous functional enrichment analysis for each random gene set followed by calculating SS values based on the enriched categories. Second, we sampled the same numbers of GO terms (14 for PSGs and 16 for convergent genes) from the GO categories of *D. melanogaster* 1000 times and calculated the SS values. For both permutations, observed SS values greater than or equal to the 95th percentiles of the random samples were considered significant.

Acclimation experiment under hypothermia or hypoxia treatment

To validate the changes in RNA expression response to the marine-montane transition, talitrids (*M. aosen*) were acclimated for 3 days in hypothermic (10°C) or hypoxic environments (the latter of which included a hypoxic gas mixture containing 13.5% oxygen, which is equivalent to an altitude of 3200 m). The talitrids fed at room temperature and normoxic environment were used as a control. The gills and legs were separately dissected for RNA extraction, and cDNA was generated using the HiScript cDNA Synthesis Kit (Vazyme). RT-qPCR was performed in a 20-μl system, with 50 ng of diluted cDNA and forward and reverse primers at 0.4 μM. The following conditions were used for PCR: 95°C for 5 min (initial denaturation), denaturation at 95°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 30 s for 30 cycles to stay within the linear range of amplification. The expression levels of key genes (*AMY1*, *ACLY*, and *OGDH*) were calculated using the $2^{-\Delta\Delta Ct}$ method. The *GAPDH* gene was used to normalize gene expression. All primers are listed in the table S27. All data were obtained from three independent assays with three replicates in each assay.

Genetic confirmation and knockout of *SMC3*

We designed primers of *SMC3* gene (*SMC3*-F, GACACGTACAACATGGAGCA; *SMC3*-R, GCAGCGTCCTCGCGTCTCTTC) for PCR amplification and sequenced across 104 species evenly selected according to phylogeny (Fig. 2A). We also implemented the "Free-ratio model" to estimate the variable dN/dS ratio on each phylogenetic branch and node based on the aligned codon sequences in PAML.

To investigate the phenotypic consequences of *SMC3*, we used a CRISPR-Cas9 strategy to knockout the *SMC3* gene in *D. magna*. The *SMC3* gene of *D. magna* has 15 exons, and the orthologous sequence of the talitrids mutation site is at exon 5. Two sgRNAs (sgRNA-1 and sgRNA-3) were designed to bind the sense strand from exons 3 and 5, respectively. One (sgRNA-2) was designed to bind the anti-sense strand from exon 4 (Fig. 4E). The target site of the *SMC3*-targeting sgRNA-1 was 5′-ATCACTTGAGACCTGAGCAAagg-3′; sgRNA-2 was 3′-ggtTCGACGAGCTCTCCTCGCGC-5′, and sgRNA-3 was 5′-GAGACAATTGCTCAAAAAGAcgg-3′. Syntheses of sgRNAs and Cas9 protein (Invitrogen) were performed as described previously (*80*).

For CRISPR somatic mutagenesis, approximately 0.2 nl of sgRNA and Cas9 mixture (400 ng/μl) were injected into embryos, which were collected from *Daphnia* within 0.5 hours after ovulation. An injected embryo was transferred into a well of a 12-well plate filled with 100 μl of 80 mM sucrose for 3 days at 22°C. The average hatching rate with the deleterious *SMC3* knockouts was 56.5%.

The phenotypes were examined under a LEICA M205C stereomicroscope. For genotyping, we extracted the DNA of single hatchlings that showed a CRISPR-induced phenotype using the TIANamp Micro DNA Kit (TIANGEN). The sgRNA-targeted genomic templates in the *SMC3* locus were amplified by PCR. The primer sequences are *SMC3*-fwd-gDNA (5′-GCGTTCGGTTGAATATACCAT-3′) and *SMC3*-rev-gDNA (5′-CACGCTGATCTCGGGATGTGA-3′). The PCR products were cloned by using the pEASY-T3 Cloning Kit (TransGen Biotech), and 20 positive clones per individual were sequenced with the M13R universal primer. *D. magna* showed the phenotypic change that was selected for analysis of expression changes by RT-qPCR. The *GAPDH* gene was used as an internal reference, and fold changes applied in the calculations are means obtained from the analyses of three biological replicates. All primers are listed in table S27.

Supplementary Materials

This PDF file includes:
Figs. S1 to S21
Legends for tables S1 and S23
Tables S2 to S22, and S24 to S27
References

Other Supplementary Material for this manuscript includes the following:
Tables S1 and S23

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. A. Meyer, S. Schloissnig, P. Franchini, K. Du, J. M. Woltering, I. Irisarri, W. Y. Wong, S. Nowoshilow, S. Kneitz, A. Kawaguchi, A. Fabrizius, P. Xiong, C. Dechaud, H. P. Spaink, J. Volff, O. Simakov, T. Burmester, E. M. Tanaka, M. Schartl, Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* **590**, 284–289 (2021).
2. X. Bi, K. Wang, L. Yang, H. Pan, H. Jiang, Q. Wei, M. Fang, H. Yu, C. Zhu, Y. Cai, Y. He, X. Gan, H. Zeng, D. Yu, Y. Zhu, H. Jiang, Q. Qiu, H. Yang, Y. E. Zhang, W. Wang, M. Zhu, S. He, G. Zhang, Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes. *Cell* **184**, 1377–1391.e14 (2021).

3. H. Glenner, P. F. Thomsen, M. B. Hebsgaard, M. V. Sørensen, E. Willerslev, The origin of insects. *Science* **314**, 1883–1884 (2006).
4. C. D. Schubart, R. Diesel, S. B. Hedges, Rapid evolution to terrestrial life in Jamaican crabs. *Nature* **393**, 363–365 (1998).
5. P. K. Strother, C. Foster, A fossil record of land plant origins from charophyte algae. *Science* **373**, 792–796 (2021).
6. C. C. Labandeira, Invasion of the continents: Cyanobacterial crusts to tree-inhabiting arthropods. *Trends Ecol. Evol.* **20**, 253–262 (2005).
7. D. Craw, P. Upton, C. P. Burridge, G. P. Wallis, J. M. Waters, Rapid biological speciation driven by tectonic evolution in New Zealand. *Nat. Geosci.* **9**, 140–144 (2015).
8. D. J. Lohmann, M. de Bruyn, T. Page, K. von Rintelen, R. Hall, P. K. L. Ng, H.-T. Shih, G. R. Carvalho, T. von Rintelen, Biogeography of the Indo-Australian archipelago. *Annu. Rev. Ecol. Evol. Syst.* **42**, 205–226 (2011).
9. J. A. Friend, A. Richardson, Biology of terrestrial amphipods. *Annu. Rev. Entomol.* **31**, 25–48 (1986).
10. T. Fujisawa, T. G. Barraclough, Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. *Syst. Biol.* **62**, 707–724 (2013).
11. W. P. Maddison, M. S. P. Otto, Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* **56**, 701–710 (2007).
12. D. L. Rabosky, M. Grundler, C. Anderson, P. Title, J. J. Shi, J. W. Brown, H. Huang, J. G. Larson, BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* **5**, 701–707 (2014).
13. P. W. Franks, F. Renström, Using genotype-based recall to estimate the effects of *AMY1* copy number variation in substrate metabolism. *Diabetes* **65**, 3240–3242 (2016).
14. J. H. F. Cavalcanti, A. Esteves-Ferreira, C. G. Quinhones, I. A. Pereira-Lima, A. Nunes-Nesi, A. R. Fernie, W. L. Araujo, Evolution and functional implications of the tricarboxylic acid cycle as revealed by phylogenetic analysis. *Genome Biol. Evol.* **6**, 2830–2848 (2014).
15. S. H. Lee, M. C. Lee, J. Puthumana, J. C. Park, S. Kang, J. Han, K. Shin, H. G. Park, A. S. Om, J. S. Lee, Effects of temperature on growth and fatty acid synthesis in the cyclopoid copepod *Paracyclopina nana*. *Fish Sci.* **83**, 725–734 (2017).
16. K. Massé, S. Bhamra, G. Allsop, N. Dale, E. A. Jones, Ectophosphodiesterase/nucleotide phosphohydrolase (Enpp) nucleotidases: Cloning, conservation and developmental restriction. *Int. J. Dev. Biol.* **54**, 181–193 (2009).
17. R. A. Mathias, T. M. Greco, A. Oberstein, H. G. Budayeva, R. Chakrabarti, E. A. Rowland, Y. Kang, T. Shenk, I. M. Cristea, Sirtuin 4 is a lipoamidase regulating pyruvate dehydrogenase complex activity. *Cell* **159**, 1615–1625 (2014).
18. J. W. Kim, C. V. Dang, Multifaceted roles of glycolytic enzymes. *Trends Biochem. Sci.* **30**, 142–150 (2005).
19. K. Mouri, S. Horiuchi, T. Uemura, Cohesin controls planar cell polarity by regulating the level of the seven-pass transmembrane cadherin Flamingo. *Genes Cells* **17**, 509–524 (2012).
20. M. Averof, S. M. Cohen, Evolutionary origin of insect wings from ancestral gills. *Nature* **385**, 627–630 (1997).
21. W. Damen, T. Saridaki, M. Averof, Diverse adaptations of an ancestral gill: A common evolutionary origin for wings, breathing organs, and spinnerets. *Curr. Biol.* **12**, 1711–1716 (2002).
22. D. Ebert, *Daphnia* as a versatile model system in ecology and evolution. *EvoDevo* **13**, 16 (2022).
23. S. Xu, Z. He, Z. Zhang, Z. Guo, W. Guo, H. Lyu, J. Li, M. Yang, Z. Du, Y. Huang, R. Zhou, C. Zhong, D. E. Boufford, M. Lerdau, C. Wu, N. C. Duke; The International Mangrove Consortium, S. Shi, The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. *Natl. Sci. Rev.* **4**, 721–734 (2017).
24. G. J. Vermeij, V. M. Watson-Zink, Terrestrialization in gastropods: Lineages, ecological constraints and comparisons with other animals. *Biol. J. Linnean Soc.* **136**, 393–404 (2022).
25. T. Maurin, C. Rangin, Structure and kinematics of the Indo-Burmese wedge: Recent and fast growth of the outer wedge. *Tectonics* **28**, TC2010 (2009).
26. G. P. Yumul, C. B. Dimalanta, V. B. Maglambayan, E. J. Marquez, Tectonic setting of a composite terrane: A review of the Philippine island arc system. *Geosci. J.* **12**, 7–17 (2008).
27. B. Stelbrink, C. Albrecht, R. Hall, T. von Rintelen, The biogeography of Sulawesi revisited: Is there evidence for a vicariant origin of taxa on Wallace's "anomalous island"? *Evolution* **66**, 2252–2271 (2012).
28. V. S. F. T. Merckx, K. P. Hendriks, K. K. Beentjes, C. B. Mennes, L. E. Becking, K. T. C. A. Peijnenburg, A. Afendy, N. Arumugam, H. de Boer, A. Biun, M. M. Buang, P.-P. Chen, A. Y. C. Chung, R. Dow, F. A. A. Feijen, H. Feijen, C. F. V. Soest, J. Geml, R. Geurts, B. Gravendeel, P. Hovenkamp, P. Imbun, I. Ipor, S. B. Janssens, M. Jocqué, H. Kappes, E. Khoo, P. Koomen, F. Lens, R. J. Majapun, L. N. Morgado, S. Neupane, N. Nieser, J. T. Pereira, H. Rahman, S. Sabran, A. Sawang, R. M. Schwallier, P.-S. Shim, H. Smit, N. Sol, M. Spait, M. Stech, F. Stokvis, J. B. Sugau, M. Suleiman, S. Sumail, D. C. Thomas, J. van Tol, F. Y. Y. Tuh, B. E. Yahya, J. Nais, R. Repin, M. Lakim, M. Schilthuizen, Evolution of endemism on a young tropical mountain. *Nature* **524**, 347–350 (2015).
29. S. A. Whattam, J. Malpas, J. R. Ali, C. H. Lo, I. Smith, Formation and emplacement of the Northland ophiolite, northern New Zealand: SW Pacific tectonic implications. *J. Geol. Soc. London* **162**, 225–241 (2005).
30. S. Maruyama, Y. Isozaki, G. Kimura, M. Terabayashi, Paleogeographic maps of the Japanese islands: Plate tectonic synthesis from 750 Ma to the present. *Isl. Arc.* **6**, 121–142 (1997).
31. M. G. Rix, M. S. Harvey, Phylogeny and historical biogeography of ancient assassin spiders (Araneae: Archaeidae) in the Australian mesic zone: Evidence for Miocene speciation within Tertiary refugia. *Mol. Phylogenet. Evol.* **62**, 375–396 (2012).
32. D. Müller, N. Flament, K. J. Matthews, S. E. Williams, M. Gurnis, Formation of Australian continental margin highlands driven by plate–mantle interaction. *Earth Planet. Sci. Lett.* **441**, 60–70 (2016).
33. C. Rahbek, M. K. Borregaard, A. Antonelli, R. K. Colwell, J. Fjelds, Building mountain biodiversity: Geological and evolutionary processes. *Science* **365**, 1114–1119 (2019).
34. L. Wu, D. Tang, C. Shen, Y. Bai, Z. Wang, Comparative transcriptome analysis of the gills of *Cardisoma armatum* provides novel insights into the terrestrial adaptive related mechanism of air exposure stress. *Genomics* **113**, 1193–1202 (2021).
35. H. S. Bruce, N. H. Patel, The *Daphnia* carapace and other novel structures evolved via the cryptic persistence of serial homologs. *Curr. Biol.* **32**, 3792–3799.e3 (2022).
36. H. S. Bruce, N. H. Patel, Knockout of crustacean leg patterning genes suggests that insect wings and body walls evolved from ancient leg segments. *Nat. Ecol. Evol.* **4**, 1703–1712 (2020).
37. Z. Hou, B. Sket, A review of Gammaridae (Crustacea: Amphipoda): The family extent, its evolutionary history, and taxonomic redefinition of genera. *Zool. J. Linn. Soc.* **176**, 323–348 (2016).
38. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 1–6 (2021).
39. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, P. A. Pevzner, SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
40. M. Seppey, M. Manni, E. M. Zdobnov, "BUSCO: Assessing genome assembly and annotation completeness" in *Gene Prediction: Methods and Protocols*, M. Kollmar, Ed. (Methods in Molecular Biology, Springer, 2019), pp. 227–245.
41. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
42. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
43. G. Moreno-Hagelsieb, K. Latimer, Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **24**, 319–324 (2008).
44. S. Kuraku, C. M. Zmasek, O. Nishimura, K. Katoh, aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res.* **41**, W22–W28 (2013).
45. R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut, A. J. Drummond, BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **15**, e1006650 (2019).
46. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
47. L. T. Nguyen, H. A. Schmidt, A. Von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
48. A. J. Aberer, K. Kobert, A. Stamatakis, ExaBayes: Massively parallel Bayesian tree inference for the whole-genome era. *Mol. Biol. Evol.* **31**, 2553–2556 (2014).
49. S. Mirarab, R. Reaz, M. S. Bayzid, T. Zimmermann, M. S. Swenson, T. Warnow, ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–i548 (2014).
50. H. Wickham, Reshaping data with the reshape package. *J. Stat. Softw.* **21**, 1–20 (2007).
51. E. L. Bousfield, G. O. Poinar, New terrestrial amphipod from tertiary amber deposits of the Dominican Republic. *J. Crustac. Biol.* **15**, 746–755 (1995).
52. C. Villacorta, D. Jaume, P. Oromi, C. Juan, Under the volcano: Phylogeography and evolution of the cave-dwelling *Palmorchestia hypogaea* (Amphipoda, Crustacea) at La Palma (Canary Islands). *BMC Biol.* **6**, 7 (2008).
53. K. Jażdżewski, J. Kupryjanowicz, One more fossil niphargid (Malacostraca: Amphipoda) from Baltic amber. *J. Crustac. Biol.* **30**, 413–416 (2010).
54. D. Ionov, Mantle structure and rifting processes in the Baikal–Mongolia region: Geophysical data and evidence from xenoliths in volcanic rocks. *Tectonophysics* **351**, 41–60 (2002).

55. S. V. Popov, F. Rögl, A. Y. Rozanov, F. F. Steininger, I. G. Shcherba, M. Kovac, Lithological-paleogeographic maps of Paratethys: 10 maps late Eocene to Pliocene. *Cour. Forsch. Senck.* **250**, 1–46 (2004).
56. T. Wilke, T. Hauffe, E. Jovanovska, A. Cvetkoska, T. Donders, K. Ekschmitt, A. Francke, J. H. Lacey, Z. Levkov, C. R. Marshall, T. A. Neubauer, D. Silvestro, B. Stelbrink, H. Vogel, C. Albrecht, J. Holtvoeth, S. Krastel, N. Leicher, M. J. Leng, K. Lindhorst, A. Masi, N. Ognjanova-Rumenova, K. Panagiotopoulos, J. M. Reed, L. Sadori, S. Tofilovska, B. Van Bocxlaer, F. Wagner-Cremer, F. P. Wesselingh, V. Wolters, G. Zanchetta, X. Zhang, B. Wagner, Deep drilling reveals massive shifts in evolutionary dynamics after formation of ancient ecosystem. *Sci. Adv.* **6**, eabb2943 (2020).
57. F. R. Schram, R. M. Feldmann, M. Copeland, The Late Devonian Palaeopalaemonidae and the earliest decapod crustaceans. *J. Paleo.* **52**, 1375–1387 (1978).
58. S. A. Smith, B. C. O'Meara, treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
59. M. Dos Reis, T. Zhu, Z. Yang, The impact of the rate prior on Bayesian estimation of divergence times with multiple loci. *Syst. Biol.* **63**, 555–565 (2014).
60. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
61. N. J. Matzke, Probabilistic historical biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* **5**, 242–248 (2013).
62. S. A. Ishikawa, A. Zhukova, W. Iwasaki, O. Gascuel, A fast likelihood method to reconstruct and visualize ancestral scenarios. *Mol. Biol. Evol.* **36**, 2069–2085 (2019).
63. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
64. L. J. Revell, Phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2011).
65. J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, Package "vegan." R Package version 2.4–1 (R Foundation for Statistical Computing, 2016).
66. R. J. Hijmans, Introduction to the 'geosphere' package (version 1.5-14). (2021).
67. T. Stadler, Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6187–6192 (2011).
68. R. G. Fitzjohn, Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* **3**, 1084–1092 (2012).
69. W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, A. O. Mooers, The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
70. F. J. Rohlf, *tpsDig, Digitize Landmarks and Outlines, Version 2.05* (Department of Ecology and Evolution, State University of New York, 2006).
71. D. C. Adams, E. Otárola-Castillo, geomorph: An R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* **4**, 393–399 (2013).
72. S. P. Blomberg, T. Garland Jr., A. R. Ives, Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).
73. W. P. Maddison, Mesquite: A modular system for evolutionary analysis. *Evolution* **62**, 1103–1118 (2008).
74. C. T. Stayton, The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution* **69**, 2140–2153 (2015).
75. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
76. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
77. Z. Zou, J. Zhang, Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol. Biol. Evol.* **32**, 2085–2096 (2015).
78. D. Bu, H. Luo, P. Huo, Z. Wang, J. Wu, S. Zhang, Z. He, Y. Wu, L. Zhao, J. Liu, J. Guo, S. Fang, W. Cao, L. Yi, Y. Zhao, L. Kong, KOBAS-i: Intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* **49**, W317–W325 (2021).
79. G. Yu, F. Li, Y. Qin, X. Bo, S. Wang, GOSemSim: An R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**, 976–978 (2010).
80. H. Kumagai, T. Nakanishi, T. Matsuura, Y. Kato, H. Watanabe, CRISPR/Cas-mediated knock-in via non-homologous end-joining in the crustacean *Daphnia magna*. *PLOS ONE* **12**, e0186112 (2017).
81. D. Nandy, Geology and tectonics of Arakan Yoma—A reappraisal. *Bull. Geol. Soc. Malays.* **20**, 137–148 (1986).
82. S. Kiel, A. G. S. Fernando, C. Y. Magtoto, T. Kase, Mollusks from Miocene hydrocarbon-seep deposits in the Ilocos-Central Luzon Basin, Luzon Island Philippines. *Acta Palaeontol. Pol.* **67**, 917–947 (2022).
83. A. M. S. Nugraha, R. Hall, Late cenozoic palaeogeography of Sulawesi, Indonesia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **490**, 191–209 (2018).
84. M. Madon, What lies beneath: Deep-marine trace fossils in the West Crocker Formation and their palaeoenvironmental significance. *War. Geol.* **45**, 346–347 (2019).
85. M. K. Eagle, Six new fossil bivalves from the early Miocene of Auckland and Northland, New Zealand. *Rec. Aust. Mus.* **36**, 141–168 (1999).
86. K. Uemura, Y. Kikuchi, H. Nagato, A. Nikaido, Middle Miocene plants from marine sediments in Tatsugoroshi of Hitachiota, Ibaraki Prefecture, Japan. *Bull. Natl. Sci. Mus. Ser. C Geol. Paleontol.* **32**, 1–11 (2006).
87. J. Pickett, M. Macphail, A. Partridge, M. Pole, Middle Miocene palaeotopography at Little Bay, near Maroubra, New South Wales. *Aust. J. Earth Sci.* **44**, 509–518 (1997).
88. K. D. Corbett, An early Miocene flathead (Pisces: Platycephalidae) from Wynyard, Tasmania. *Pap. Proc. R. Soc. Tasmania* **114**, 165–175 (1980).

Acknowledgments: We thank D. K. Kurenshchikov (Russia), O. Alfaruq (Indonesia), M. H. Aung (Myanmar), S.-T. Kim (Korea), Z. Zhao, H. Zhao, and Q. Zhao for assistance in field collection; A. Ugolini and L. Xie for specimen donation; P. Zhang for help in the production of phylogenomic markers; M. Bai for assistance in morphological studies; Y. Hu, J. Fu, H. Qiao, W. Zhang, L. Zhang, B. Guo, and H. She for helpful discussions. **Funding:** This study was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (grant numbers XDB31000000 and XDA2005020102), National Natural Science Foundation of China (grant number 32070423), and the Third Xinjiang Scientific Expedition Program (grant number 2021xjkk0600). **Author contributions:** H.L., S.L., and Z.H. designed research. H.L., Y.Z., B.Z., Y.T., W.X., and M.H. performed research. Y.H., W.C., F.B., and S.L. collected samples. H.Y. and P.J. participated in discussions and provided suggestions. H.L. and Z.H. wrote the manuscript. All authors proofread and approved the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All the Sanger sequences are deposited in GenBank and the accession number (OQ475018-OQ475940, OQ512158-OQ512718, OQ514061-OQ514290, and OQ514350-OQ515447) can be found in table S1, raw RNA-seq, and genome data generated in this study are deposited into the NCBI SRA under the accession number PRJNA938803. The alignment files, phylogenetic trees, morphological data, and other data used in this study are publicly available on https://doi.org/10.57760/sciencedb.07691.

Submitted 21 December 2022
Accepted 17 May 2023
Published 23 June 2023
10.1126/sciadv.adg4011