nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-018-0494-6

Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary

Michael E. Alfaro[1,10*], Brant C. Faircloth[2,3,10*], Richard C. Harrington[4,5], Laurie Sorenson[1], Matt Friedman[4,6,7], Christine E. Thacker[8], Carl H. Oliveros[2], David Černý[1] and Thomas J. Near[5,9]

The Cretaceous–Palaeogene (K–Pg) mass extinction is linked to the rapid emergence of ecologically divergent higher taxa (for example, families and orders) across terrestrial vertebrates, but its impact on the diversification of marine vertebrates is less clear. Spiny-rayed fishes (Acanthomorpha) provide an ideal system for exploring the effects of the K–Pg on fish diversification, yet despite decades of morphological and molecular phylogenetic efforts, resolution of both early diverging lineages and enormously diverse subclades remains problematic. Recent multilocus studies have provided the first resolved phylogenetic backbone for acanthomorphs and suggested novel relationships among major lineages. However, these new relationships and associated timescales have not been interrogated using phylogenomic approaches. Here, we use targeted enrichment of >1,000 ultraconserved elements in conjunction with a divergence time analysis to resolve relationships among 120 major acanthomorph lineages and provide a new timescale for acanthomorph radiation. Our results include a well-supported topology that strongly resolves relationships along the acanthomorph backbone and the recovery of several new relationships within six major percomorph subclades. Divergence time analyses also reveal that crown ages for five of these subclades, and for the bulk of the species diversity in the sixth, coincide with the K–Pg boundary, with divergences between anatomically and ecologically distinctive suprafamilial clades concentrated in the first 10 million years of the Cenozoic.

The mass extinction event at the Cretaceous–Palaeogene (K–Pg) boundary has been linked to the rapid appearance of anatomically and ecologically distinctive higher-level taxa in major clades of terrestrial vertebrates[1,2], including birds[3,4], placental mammals[5] and squamates[6]. Surprisingly, although the K–Pg mass extinction exerted a disproportionate effect on marine taxa[7], molecular phylogenetic timescales do not identify the K–Pg as an important factor underlying present-day diversity in marine fishes[8–10]. Instead, these studies suggest that the origins of modern marine fish diversity were established through a series of radiations over the Late Cretaceous and early Cenozoic[8,11]. Although the fossil record suggests that family-level fish diversity was not significantly impacted by the K–Pg mass extinction[12–14], significant phenotypic evolution[15,16] and restructuring of marine fish communities[17] occurred in the wake of this event. One of the most striking patterns to emerge from the fossil record, but not molecular timescales, is a sharp rise in the diversity of families of acanthomorph fishes in the early Cenozoic[15,16,18].

Acanthomorphs are the dominant clade of aquatic vertebrates, comprising more than 18,100 species including icons of adaptive radiation such as cichlids and Antarctic notothenioids, major stocks of marine commercial fisheries such as tunas, jacks, flatfishes and rockfishes, and families associated with coral reef communities worldwide, including butterflyfishes, angelfishes and wrasses[8,19]. The acanthomorph fossil record extends back nearly 100 Myr[18]. Despite the astonishing diversity of this clade, the factors contributing to its evolutionary success remain poorly understood. Resolving evolutionary relationships within acanthomorphs represents a key historical challenge to understanding the emergence of the remarkable variety within the clade[18,20,21]. This is most conspicuous for percomorphs—the acanthomorph equivalent of placental mammals and neoavian birds. Like those terrestrial groups, percomorphs display extreme morphological and taxonomic diversity in the modern fauna, represent an exceptional radiation in terms of rates of lineage diversification relative to other vertebrate clades[22], show an abrupt rise in the appearance of new lineages and bodyplans in the early Palaeogene fossil record[15], and have long vexed systematists aiming to dissect their major patterns of intrarelationships[20].

The seeming intractability of a well-resolved phylogenetic framework for the ~17,190 percomorph species classified into 264 recognized families led to this clade being referred to as the 'bush at the top' of the acanthomorph tree[23]. More recently, multilocus molecular studies have identified several novel percomorph subclades and a working phylogenetic backbone connecting them[8,9,11,24] (Figs. 1 and 2). For example, Ovalentaria[25] comprises a hyperdiverse assemblage including cichlids, silversides, needlefishes, killifishes, damselfishes and blennioids; tunas and their allies form a surprising sister group to a clade including seahorses, pipefishes, goatfishes and flying gurnards[8,24]; and flatfishes form the sister group to jacks, billfishes and their allies[8,24]. This resolution of the evolutionary history of percomorphs is exciting and unprecedented. However, the delineation of recently resolved groups rests largely on the analysis of datasets relying on a relatively small set of 'workhorse' legacy markers and remains largely untested by genome-scale analyses[26]. Furthermore, the backbones of these recently published phylogenies reveal regions with conflicting resolutions, short internodes

[1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA. [2]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA. [3]Museum of Natural Science, Louisiana State University, Baton Rouge, LA, USA. [4]Department of Earth Sciences, University of Oxford, Oxford, UK. [5]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. [6]Museum of Paleontology, University of Michigan, Ann Arbor, MI, USA. [7]Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA. [8]Natural History Museum of Los Angeles County, Los Angeles, CA, USA. [9]Peabody Museum of Natural History, Yale University, New Haven, CT, USA. [10]These authors contributed equally: Michael E. Alfaro and Brant C. Faircloth. *e-mail: michaelalfaro@ucla.edu; brant@faircloth-lab.org

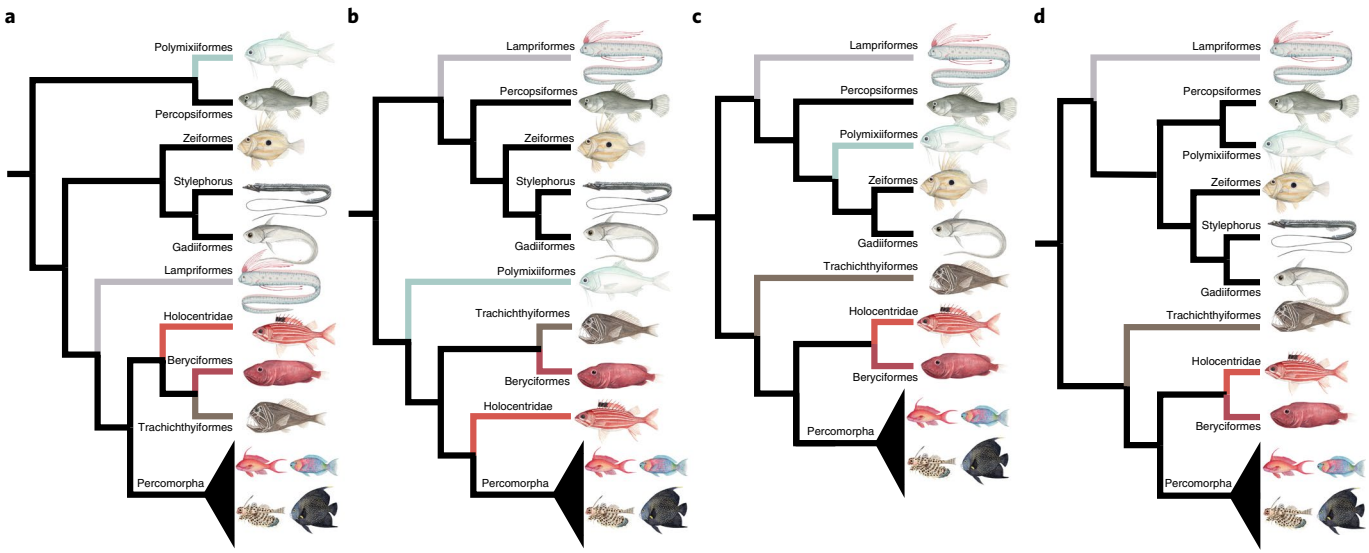

Fig. 1 | Previous hypotheses for relationships among early diverging acanthomorphs. Recent multilocus studies conflict over the position of lampriforms, polymixiiforms and the arrangement of 'beryciform' lineages, which are thought to be most closely related to the percomorphs. **a**, Ref. [8] resolves polymixiiforms as the sister group to percopsiforms, and lampriforms as sister to a clade comprising holocentrids + beryciforms (here taken to include berycoids and stephanoberycoids) + trachichthyiforms and percomorphs. **b**, Ref. [24] resolves lampriforms as the sister lineage to a clade containing percopsiforms, zeiforms and gadiiforms, and places polymixiiforms as the sister lineage to a clade with beryciforms + trachichthyiforms and holocentrids as successive sister lineages to percomorphs. **c**, Ref. [60] resolves lampriforms as the sister group to percopsiforms, polymixiiforms and zeiforms + gadiiforms, and resolves trachichthyiforms as the sister group to a clade consisting of beryciforms + holocentrids and percomorphs. **d**, The UCE topology resolves the backbone of the acanthomorph tree with high support, revealing an early divergence between a clade containing lampriforms, percopsiforms, polymixiiforms, zeiforms and gadiiforms, and a clade containing trachichthyiforms as the sister lineage to a clade consisting of beryciforms + holocentrids and percomorphs.

and weak support, limiting our ability to investigate evolutionary scenarios for acanthomorph diversity.

Here, we present a phylogenetic analysis of 118 acanthomorph fishes and two outgroups based on DNA ultraconserved elements (UCEs)[27,28]. Our taxonomic sampling includes 77 families representing major lineages across the acanthomorph radiation and includes multiple species from percomorph clades recently proposed by other molecular phylogenetic studies. In addition, we sample representative lineages of all known early diverging acanthomorphs, including opahs, ribbonfishes, oarfishes, troutperches and beardfishes, because recent molecular studies have produced conflicting relationships among these taxa[8,24]. Expanding on the sequencing strategy of earlier genome-scale efforts for fishes[26], we used a targeted enrichment approach to collect data from >1,000 UCE loci. Our primary analyses were based on a data matrix that was 75% complete, included trimmed alignments that had an average length of 302 base pairs (bp) (range: 80–749 bp), and included 302,488 characters and 115,622 parsimony informative sites (Supplementary Information).

Results

Bayesian analysis of the concatenated matrix produced a fully resolved tree with high posterior support (≥0.95) for all nodes (Fig. 3) except for the sister group relationship between a clade comprising dolphinfishes, cobia and jacks, and a clade comprising archerfishes, the moonfish and billfishes (node 191). The UCE topology provides important new insight into the controversy over relationships among the earliest diverging acanthomorphs (Fig. 1) by resolving these lineages into a single clade (node 120). Specifically, lampriformes are the sister group of a clade containing two lineages: one comprising beardfishes (*Polymixia*) and troutperches + pirate perches, and the other containing dories (Zeiformes), cod (Gadiformes) and the tube-eye (*Stylephorus chordatus*). This arrangement contrasts with those proposed in foundational morphological studies[20], but shows a close correspondence with the results of anatomical analyses incorporating fossils along with living taxa, disagreeing principally in relation to the position of the root of the acanthomorph tree[21].

Our UCE-inferred phylogeny resolves percomorphs into eight major lineages (Figs. 2 and 3). The eel-like ophidiiforms and benthic batrachoidiforms are successive sister groups to all remaining percomorphs, reflecting a consistent pattern in acanthomorph molecular phylogenetics[8,24]. The gobiarians (nurseryfishes, apogonids and gobies; Fig. 3, node 139) are resolved as sister to all other non-ophidiiform and batrachoidiform percomorph lineages, with strong support throughout the goby and apogonid lineages. Seahorses, dragonfishes and pipefishes are united with the phenotypically disparate dragonets, flying gurnards and goatfishes to form Syngnatharia (node 155). The pelagiarians (tunas and their allies, node 160) are resolved as the sister group to the syngnatharians in a key point of congruence with recent multilocus studies[8,24]. The UCE topology resolves eupercarians (node 200) as a clade, which is a hyperdiverse lineage that includes basslets, groupers, centrarchids, tetraodontiforms and wrasses. Eupercaria forms the sister group to two remaining percomorph lineages: ovalentarians (cichlids, needlefishes, damselfishes and blennies; node 173) and carangarians (jacks, billfishes, flatfishes, barracudas and snooks; node 183). Notably, the UCE-inferred phylogeny provides strong node support for many novel relationships within these major clades that is largely robust to analytical method and partitioning scheme (Supplementary Information).

We inferred a timescale for the acanthomorph radiation by constraining ages for the root and 12 internal nodes spanning divergences across the tree based on robustly supported fossil assignments (Fig. 4). Calibrations comprise minima based on specific fossil specimens, combined with empirically informed 'soft' upper bounds on age priors based on sequences of outgroup ages, drawing on 12 additional, phylogenetically constrained

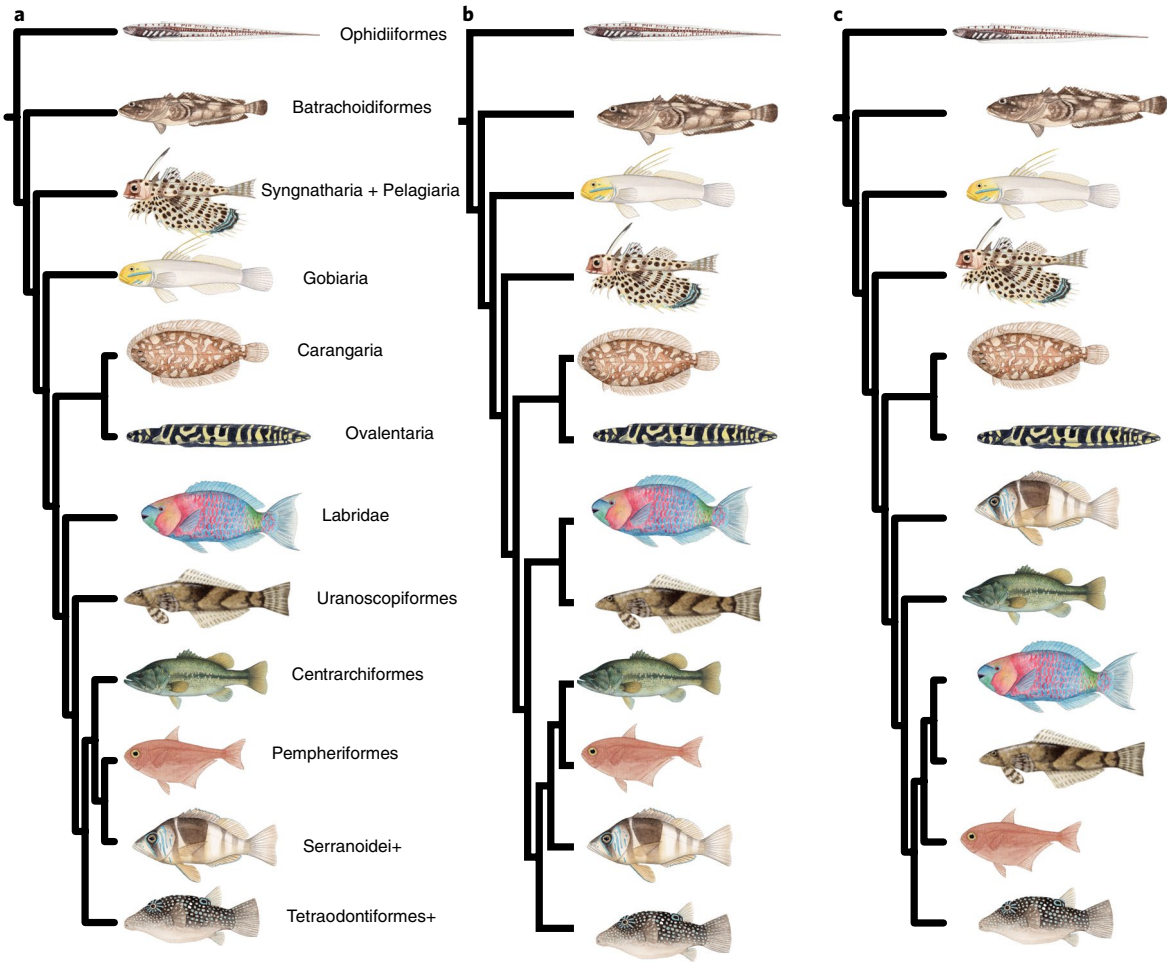

Fig. 2 | Previous hypotheses for relationships among major percomorph lineages. Previous broad-scale multilocus studies recover conflicting relationships among major lineages of percomorphs. **a**, Relationships among 12 major percomorph lineages described by ref. [8]. **b**, Ref. [24] presents a topology with multiple points of disagreement among these lineages including: (1) the relative position of gobiarians and syngnatharians + pelagiarians; (2) the relationship between labrids and other percomorph lineages; (3) the sister group of the centrarchiforms; and (4) the position of Serranoidei and their cousins. Many of these backbone relationships receive ambiguous support in ref. [8] and ref. [24] **c**, The UCE topology provides novel resolution of the major lineages identified by ref. [8], including a sister group relationship between gobiarians + syngnatharians + pelagiarians and all other percomorphs (minus ophidiiforms and batrachoidiforms), a sister group relationship between labrids and uranoscopiforms, pempheriforms as the sister group to labrids + uranoscopiforms, and an arrangement of eupercarian lineages that includes an expanded serranoid clade (Serranoidei+), centrarchiformes, a highly disparate clade including tetraodontiforms, chaetodontids and acanthuroids (Tetraodontiformes+), and the pempheriforms as successive sister lineages to a clade containing labrids and uranoscopiforms.

acanthomorph fossils and 10 additional fossils of non-acanthomorphs (Supplementary Information). Our analysis reveals two major results: (1) divergences among major percomorph lineages excluding the deeply diverging ophidiiforms and batrachoidiforms occurred in a 15 Myr window between approximately 85 and 100 million years ago (Ma) in the Late Cretaceous; and (2) remarkable congruence of the origin of five of the six principal crown groups (Syngnatharia, Pelagiaria, Eupercaria, Ovalentaria and Carangaria) close to one another as well as the K–Pg boundary (Figs. 3 and 5). Furthermore, early divergences within each of these five groups occur at or near the K–Pg boundary. The sixth group (Gobiaria) has a crown age of ~89 Myr (Fig. 3, node 139), placing its origin firmly within the Late Cretaceous. However, within Gobiaria, the node subtending more than 99% of the 2,000 living species of Gobioidei is estimated at ~66 Myr (Fig. 3, node 145), aligning the timing of origin for this clade with the other crown groups. The concordance in divergence times across major acanthomorph lineages is robust to the inclusion of contentious Cretaceous fossils, often interpreted as tetraodontiforms, in the calculation of upper bounds on calibration priors (Supplementary Information); incorporation of these records results in slightly older mean ages for most crown groups, but does not change the overall interpretation of our results (Supplementary Fig. 7).

Discussion

The pulsed origin of higher-level percomorph lineages in close proximity to the K–Pg contrasts with earlier molecular phylogenetic timescales that suggested a gradual accumulation of major clades over the Late Cretaceous and early Cenozoic[8,11], as well as whole mitochondrial genome studies that suggest more ancient divergences among acanthomorphs[29,30]. Compared with these earlier studies, we have included additional fossil constraints, consistently employed an outgroup-informed approach to the design of calibration upper bounds (Supplementary Information) and added substantially more sequence data, which should improve branch length estimation[31]. Moreover, mitochondrial data are known to systematically overestimate divergence times[32], possibly due to substitution saturation[32], lack of recombination[33] or greater susceptibility to population size

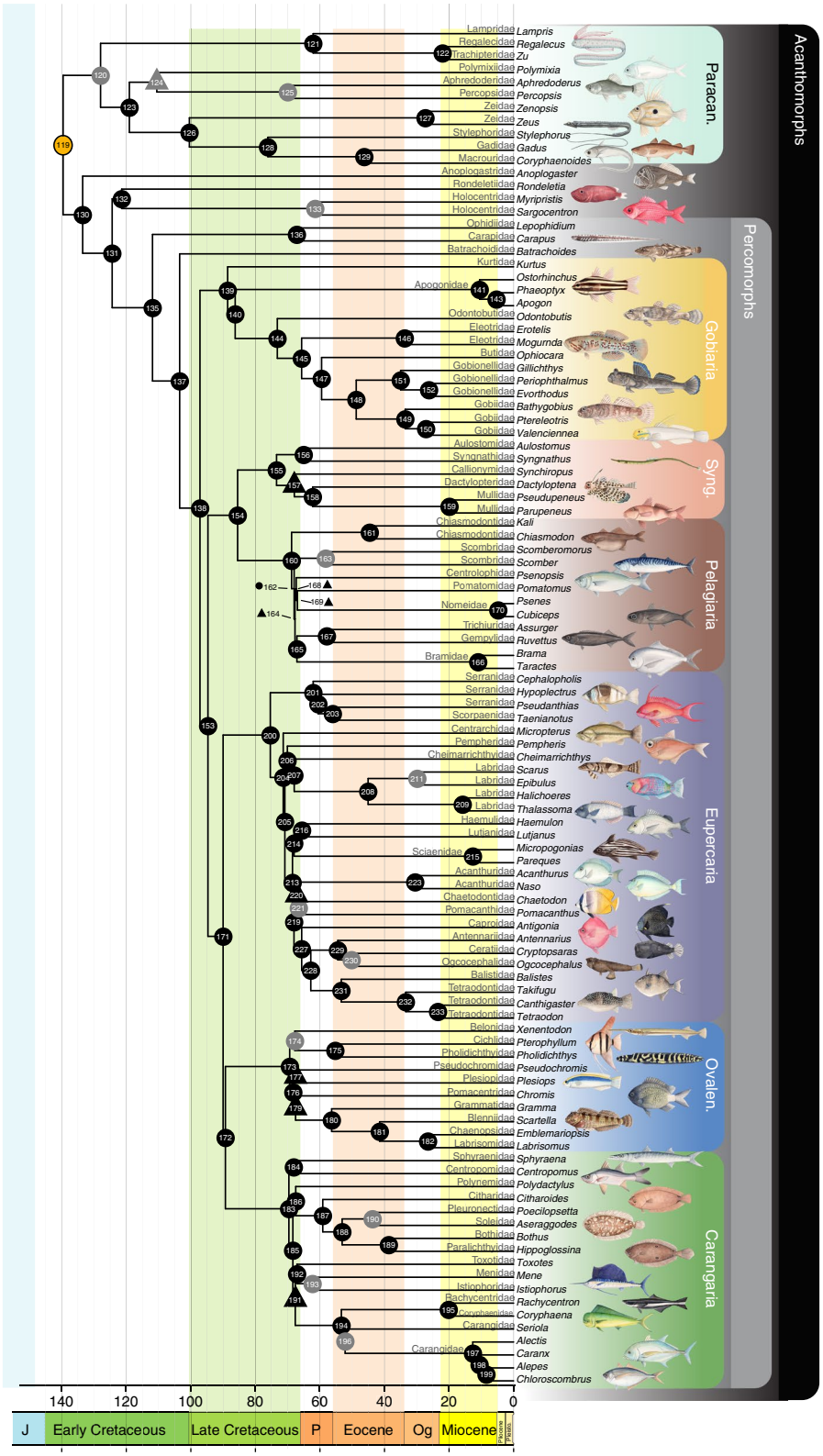

Fig. 3 | Evolutionary timescale for acanthomorph fishes. Phylogeny inferred for 120 species of acanthomorph fishes based on Bayesian analysis of 1,100 UCE loci in the partitioned, 75% complete matrix using ExaBayes[47]. All nodes in the phylogeny are supported by a posterior probability ≥0.95 except node 191 (posterior probability = 0.82). Nodes with non-parametric, maximum-likelihood bootstrap support ≥70% are indicated by circles. Nodes with bootstrap support <70% are indicated by triangles. Grey nodes represent fossil-constrained ages, whereas the orange node indicates the constraint placed on the root. Tip labels indicate the genus of sampled species, with family designations in grey text to the left. Shaded tabs indicate acanthomorph higher taxonomic groups, including six major percomorph subdivisions. The time axis represents divergences in Myr. See Supplementary Table 1 for details of taxonomic samples and Supplementary Information for additional phylogenetic analyses. J, Jurassic; Og, Oligocene; Ovalen., Ovalentaria; P, Palaeocene; Paracan., Paracanthopterygii; Pleisto., Pleistocene; Syng., Syngnatharia.

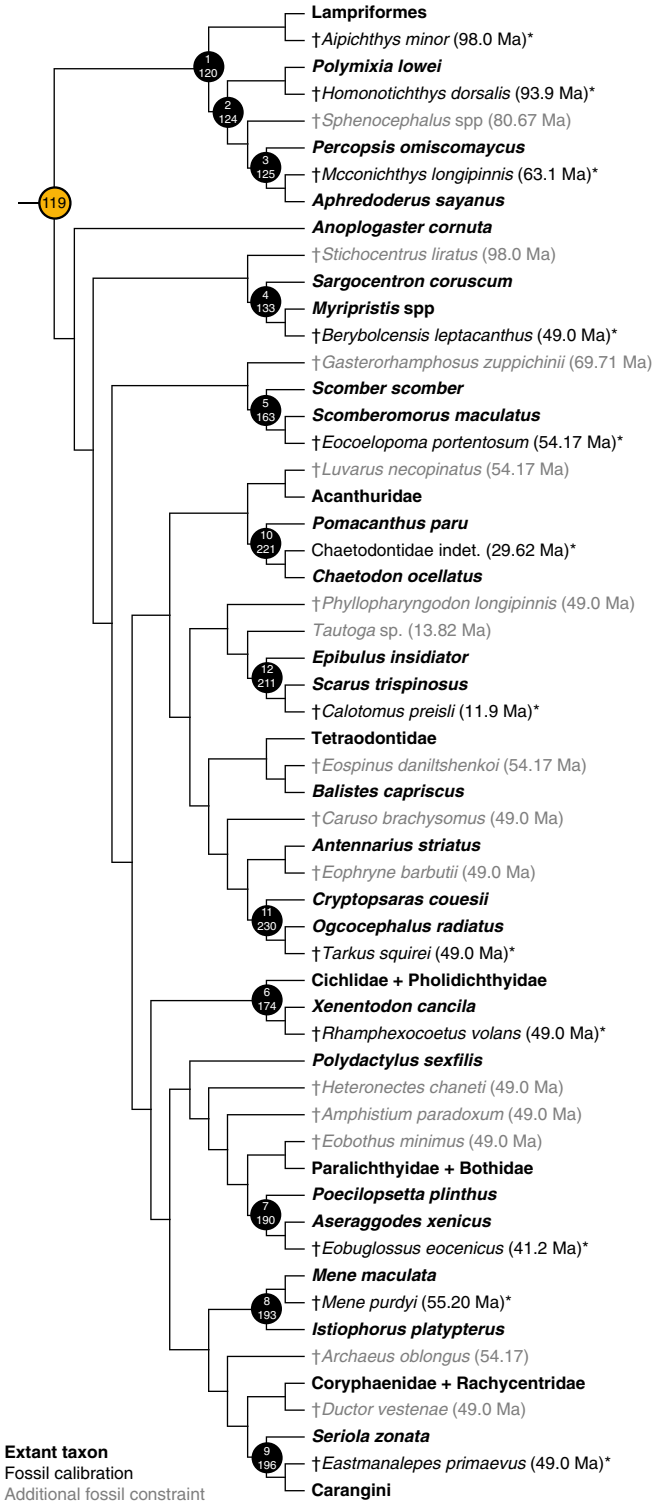

Fig. 4 | Fossil calibration placement. Skeleton of complete UCE phylogeny presented in Fig. 3, showing the placement of fossils representing calibration minima (fine black type) and those providing additional constraints using the outgroup-based approach to the estimation of 95% oldest bounds on calibration priors (fine grey type). The orange node indicates the constraint placed on the root. The top number within each dark circle represents the calibration point described in the Supplementary Information, and the bottom number corresponds to the labelled nodes in Fig. 3. The dagger symbol indicates extinct taxa. An asterisk specifies taxa used as primary node calibrations, as opposed to fossil taxa without an asterisk and shown in grey that represent outgroup sequences used to inform oldest bounds on calibration priors. For full justification of phylogenetic placement and age assessment of fossils, see the Supplementary Information.

changes[34], and the transition from mitogenomic to multilocus nuclear timescales increases the concordance of molecular dating estimates with the known fossil record for other major vertebrate radiations[3,4,33].

Our phylogenomic timescale is consistent with two conspicuous patterns in the early Palaeogene fossil record: the rise of percomorph familial diversity[15,16] and a massive change in marine

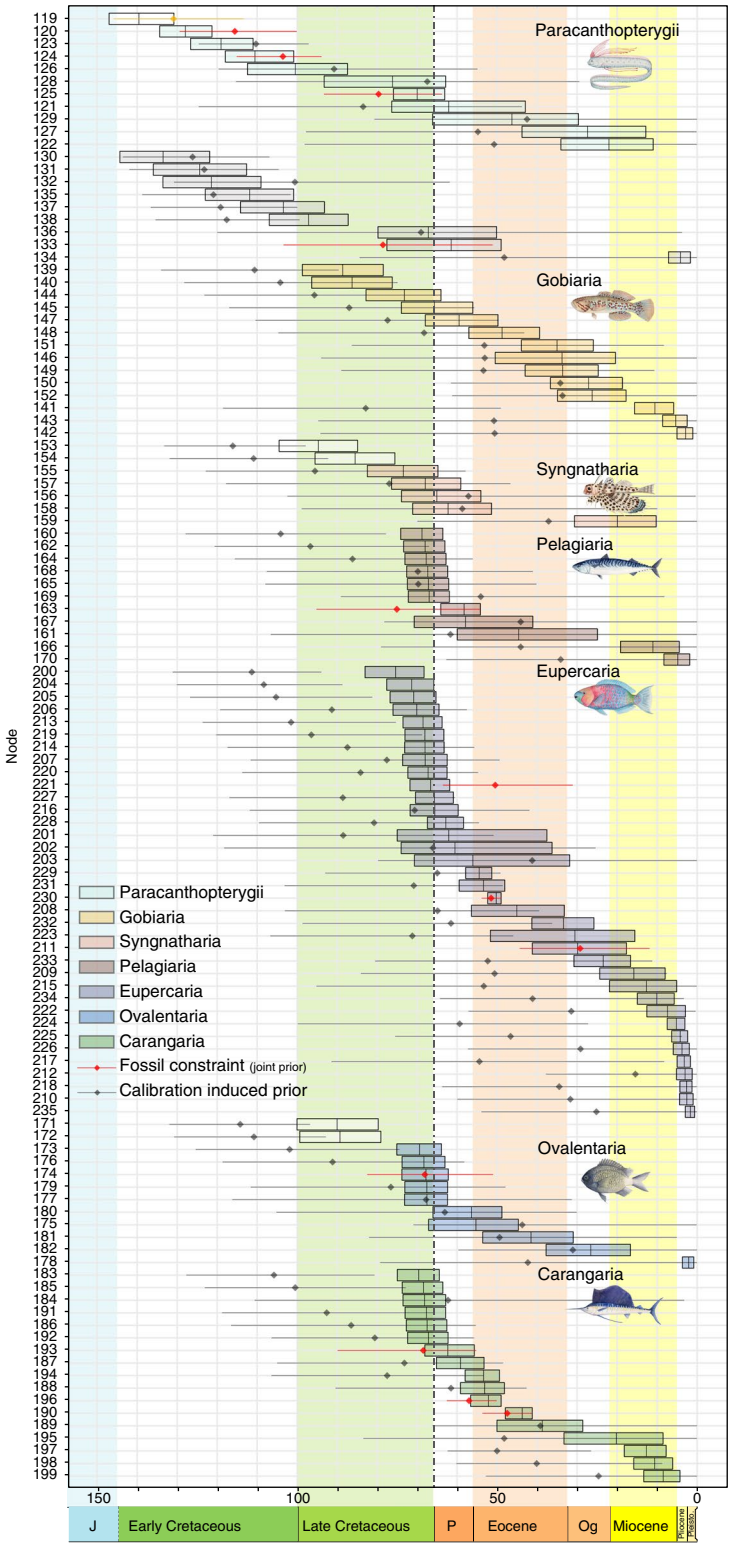

Fig. 5 | Divergence times of the acanthomorph radiation. Divergence time estimation was performed using MCMCTree[54] with fossil constraints on the root and 12 internal nodes. The node numbers correspond to the labelled nodes in Fig. 3. The fossil calibration-induced prior node age range is indicated by lines (small circles indicate the mean prior age, whereas red and orange lines and circles indicate joint prior ranges of fossil-constrained nodes). Rectangles indicate each node's 95% posterior age range (mean age estimate indicated by vertical lines). Rectangle colours correspond to the major percomorph subclades indicated in Fig. 3. We provide justification for the calibration points and additional details regarding MCMCTree analyses in the Supplementary Information. J, Jurassic; Og, Oligocene; P, Palaeocene; Pleisto., Pleistocene.

ecosystems following the K–Pg event[17]. Past studies have suggested that these two aspects are strongly linked, with the extinction of dominant Mesozoic fish clades[16,35] catalysing the opportunistic radiation of percomophs into emptied regions of ecospace[15,36,37]. Our phylogeny and associated age estimates show broad agreement with the acanthomorph fossil record, although we acknowledge that

future phylogenomic studies with more extensive sampling of modern lineages and fossil calibrations will be vital to testing this revised timeline. The earliest fossil acanthomorphs, all of mid-Cretaceous age, include members of the earliest diverging lineages of spiny-rayed fishes: lampriformes, polymixiiforms and beryciforms[18,21]. Plectocretacicoids aside, the earliest definitive body fossils of crown percomorphs derive from near the Campanian–Maastrichtian boundary, roughly 6 Myr before the K–Pg event. Consistent with our timescale and topology, these rare taxa are linked to early diverging percomorph lineages: ophidiiforms, syngnatharians and potentially batrachoidiforms[38]. In contrast, the latest Palaeocene and early Eocene deposits yield a great diversity of taxa from phylogenetically derived percomorph clades. Similar patterns are apparent in the denser otolith record, which yields a parallel archive to that provided by more sporadically distributed body fossils[39]. The picture of percomorph evolution emerging from palaeontology thus shows close correspondence to the more established cases of birds and placental mammals: rare, fragmentary, or controversial examples of deeply branching lineages in the Late Cretaceous, followed by the sudden appearance of ecologically and anatomically disparate clades in the early Palaeogene[3–5].

Conclusions

The five principal percomorph groups linked to the K–Pg (as well as the gobies) constitute an enormous range of body shapes, sizes and life histories, and capture nearly the entirety of ecological diversity explored by shallow-water and pelagic marine fishes. Our phylogenomic timescale suggests that this stunning richness derives from a series of adaptive radiations in the wake of a mass extinction, and implicates the K–Pg as the primary driver of the modern 'age of fishes'[8,17]. Given the spectacular ecological diversity, spatial range and time spanned by these groups, a full understanding of the proximate causes for diversification within them as well as the contingencies that limited diversification in percomorph lineages surviving the K–Pg (including some gobarians) requires further study. However, our results reveal the K–Pg event, and subsequent recovery of affected lineages, as a transformative episode across major lineages of terrestrial and aquatic vertebrates.

Methods

Laboratory. We obtained tissues for 118 acanthomorph species spanning 76 families (University of California Los Angeles Institutional Animal Care and Use Committee number: 2008-176-21; Supplementary Table 1), including most major lineages identified in recent multilocus studies[8,24], and extracted DNA from these species following the DNeasy protocol for extraction (Qiagen). We prepared single-[40] or dual-indexed[41] DNA libraries for targeted enrichment following slightly different library preparation and enrichment procedures for particular samples (Supplementary Table 1). We used a custom bait set targeting UCE loci identified in acanthomorph fishes to enrich UCEs from pooled sequencing libraries following a standardized protocol (versions 1.4 and 1.5 available from http://ultraconserved.org) modified from refs [26,42]. We sequenced enriched libraries using several runs of Illumina sequencing on either the MiSeq or the HiSeq platform with different target read lengths (Supplementary Table 1). A full description of the laboratory methods is available in the Supplementary Information.

Data processing. We assembled trimmed FASTQ data into contigs using a parallel wrapper around Trinity version r2013-02-25 (ref. [43]) from the PHYLUCE[44] package and used additional scripts within PHYLUCE to identify UCE loci from assembled contigs, align UCE loci across sampled lineages, trim the resulting alignments, compute alignment statistics and prepare alignments for phylogenetic analysis. We used these steps to create two subsets of alignments that were 75% complete (90 of 120 taxa have data present in each alignment) and 95% complete (114 of 120 taxa have data present in each alignment). A full description of the data-processing methods is available in the Supplementary Information.

Phylogenetic analyses. *Data partitioning.* We concatenated individual loci in the 75 or 95% subsets to different PHYLIP-formatted supermatrices (format_nexus_files_for_raxml.py) and used the concatenated 'charset' data output by this programme to construct an input file for PartitionFinder[45], which we used to partition the data with the hcluster[46] search scheme, equal weighting for overall rates, base frequencies, model parameters and the alpha parameter, and model selection by Bayesian information criterion. PartitionFinder yielded 119 partitions for the 75% complete matrix and 35 partitions for the 95% complete matrix.

Concatenated analyses. We inferred a Bayesian phylogeny using the MPI version of ExaBayes version 1.4.1 (ref. [47]). For each dataset (75% complete matrix, partitioned; 75% complete matrix, unpartitioned; 95% complete matrix, partitioned; and 95% complete matrix, unpartitioned), we ran four independent runs of 1.75 M iterations each (burnin: 25%; thinning = 500) on 4 16 CPU, 32 GB random access memory high-performance computing nodes (mike.hpc.lsu.edu). For the partitioned Bayesian analyses, we used three heated chains to sample the posterior distribution because the traces from preliminary runs suggested that some runs were becoming stuck in local optima. We visually assessed runs for convergence in Tracer (http://beast.bio.ed.ac.uk/tracer) by checking for effective sample sizes that were >200 for all model parameters and calculating the standard deviation of the split frequencies using the postProcParams and sdsf programmes included with the ExaBayes package. We also visually examined traces and effective sample size values for estimated parameters using Tracer. We created a consensus tree from the independent runs using the consense programme from the ExaBayes package (Fig. 3).

Using the same partitioning schemes suggested by PartitionFinder, we analysed the concatenated 75 and 95% data matrices using RAxML 8.0.19 (ref. [48]). We used the RAxML PTHREADS binary with the GTRGAMMA site rate substitution model on single 12 CPU, 24 GB random access memory nodes to conduct 32 maximum-likelihood searches for the phylogenetic tree that best fit the data. Following the best tree search, we used RAxML to generate non-parametric bootstrap replicates using the autoMRE option. Following the inference of the best tree and bootstrap replicates, we reconciled the best-fitting maximum-likelihood tree with the bootstrap replicates using RAxML. We compared the resulting trees with the Bayesian topology (Supplementary Figs. 1–3). The backbone of the 75% maximum-likelihood topology was identical to the Bayesian topology, although relationships within several acanthomorph subclades showed minor differences (Supplementary Fig. 3).

To account for possible artefacts introduced to downstream phylogenetic inference by base compositional differences among loci[49,50], we coded all nucleotide positions in the 75% concatenated dataset by converting adenine + guanine to 0 (R) and cytosine + thymine to 1 (Y). We analysed this unpartitioned, RY-coded matrix using RAxML, as described above, except that we used the BINGAMMA site rate substitution model in place of the GTRGAMMA model (Supplementary Fig. 4).

Coalescent-based species tree analysis. To account for coalescent stochasticity among individual UCE loci and to address the related problem where concatenated analyses can return highly supported but incorrect trees, we inferred a species tree from individual gene trees using a two-step process. First, we used PHYLUCE to resample the 75% complete matrix by loci and by sites[51] and create 100 bootstrapped subsets of our 75% complete dataset. Then, we used ASTRAL-II[52] to infer species trees from each of the bootstrapped subsets of loci, generated a majority-rule consensus tree[53] of the results (minimum clade frequency = 0.7) and compared that tree with the Bayesian analysis of the 75% complete, partitioned data matrix (Supplementary Fig. 5). It is important to note that while ASTRAL is not strictly a coalescent method, it is statistically consistent with the multispecies coalescent model and scales well to larger numbers of loci[52].

Divergence time analysis. We used MCMCTree from the PAML package[54] to estimate divergence times using the Bayesian consensus tree based on the partitioned, 75% complete matrix inferred using ExaBayes (Fig. 1), the 95% complete matrix of sequence data, a constraint on the root and 12 additional nodes using well-justified fossil placements spanning major acanthomorph lineages at shallow and deep divergences (Fig. 4). We provide full justification for fossil placements, minimum ages and outgroup sequences in the Supplementary Information.

Before running our divergence time analyses, we conducted a number of test runs with MCMCTree and our fossil calibrations. These analyses exhibited a variety of convergence problems that did not appear to improve with increasing the number of generations in the MCMC analysis. Based on the results from these test runs and similar results from other analyses, we identified two issues that appeared to contribute to the lack of convergence: the amount of missing data in the 75% complete matrix and rate differences across variable sites spanning conserved and flanking regions of UCE loci[28]. To account for the rate differences among UCE sites while also producing a partitioning scheme that could be implemented in MCMCTree, we rooted the tree inferred from the ExaBayes analysis of the 75% complete data matrix on *Alepisaurus ferox*, dropped the tip representing the outgroup lineages (*A. ferox* and *Ceratoscopelus warmingii*) and input the resulting 95% complete supermatrix and 75% complete guide tree to DendroPy[53], where we calculated the number of parsimony informative changes in each site pattern. We then input the distribution of parsimony informative changes in each site pattern to R, where we plotted the within-group sum of squares of parsimony score by a number of potential clusters (range: 1–15) using the kmeans library. Visual examination of the resulting plot suggested that four was the appropriate number of clusters (Supplementary Fig. 6). We used kmeans to compute the cluster

means and assigned individual sites to one of the four clusters. We then used a programme (get_aligns_partitioned_by_cluster.py) in PHYLUCE[44] to create a partitioned, concatenated alignment from the 95% complete matrix of sequence data that was suitable for input to PAML.

We estimated a mean substitution rate for the entire UCE alignment using BASEML (a programme in the PAML[54] package) with a strict molecular clock and the Bayesian consensus topology derived from ExaBayes analysis of the 75% complete matrix, and we used this substitution rate estimate to inform the regene_gamma prior in MCMCTree. We then followed the two-step procedure outlined in ref. [55] to infer divergence times under an approximation of the likelihood function using MCMCTree, our Bayesian consensus topology from the 75% complete matrix and the 95% complete concatenated alignment that we partitioned into four rate categories (ndata = 4), assigning an HKY + Gamma model to each partition with five gamma categories and the following parameter values for the control file: alpha = 0.1, ncatG = 5, BDparas = 1 1 0, kappa_gamma = 6 2, alpha_gamma = 1 1, regene_gamma = 2 203.72 1, sigma2_gamma = 2 5 1 and finetune = 1 (all BASEML and MCMCTree control files are available from Dryad). We performed 10 separate analyses under a model of independent rates with a burnin of 5,000 and a sample frequency of 250. We collected 10,000 samples from each run and used Tracer and LogCombiner (http://beast.bio.ed.ac.uk/tracer) to visually confirm that each run was converging on the same posterior estimates for model parameters and combine samples across runs. The effective sample size for all parameters in the combined pool was ≥200. The results from these dating analyses are presented in Figs. 3 and 5, and we present additional details in the Supplementary Information, showing the effects of including the plectocretacicoids in our analyses (Supplementary Fig. 7).

Data filtering. UCEs typically exhibit high levels of rate heterogeneity across core and flank regions, as well as within the flanking regions themselves. To test the sensitivity of our results to the sites included in the analysis, we performed a multistep, locus-filtering procedure based on the SortaDate[56] workflow. To create data partitions with more homogenous rates that could be better fit with models of sequence of evolution, we divided the multiple sequence alignment of each locus from the 75% complete matrix into 50-bp chunks using a combination of a publicly available Python script (https://gist.github.com/anonymous/ee514997912f62392a5964ca457cabf6) and a custom R script (available from Dryad). When the length of a locus was not a multiple of 50, the remainder was retained as a separate chunk.

We then inferred the best maximum-likelihood tree from each of the 6,543 resulting chunks using the 'GTRGAMMA' substitution and site heterogeneity model in RAxML 8.1.20 (ref. [48]), as implemented in the ETE 3 toolkit[57], on an Intel Core i7, 16 GB random access memory machine. We also computed SH-like support values for each best tree using the '-f J' option, and RAxML inferred trees with support values in 4,237 cases, corresponding to a success rate of 64.8%. The failures to infer a tree were due to invariant or nearly invariant chunks (due to their short lengths) or the presence of one or more taxa with no data in the sequence matrix.

Using a custom Python script (available from Dryad), we collapsed all nodes in the remaining RAxML chunk trees that had SH-like support values of 0.9 or lower. We then passed trees with collapsed nodes to the SortaDate package[56], which we used to compute the root-to-tip branch length variance, tree length and proportion of bipartitions shared with the Bayesian consensus tree for each of the 4,237 chunk trees, with *A. ferox* and *C. warmingii* as the outgroups. Because some input trees were missing one or both outgroup taxa, some did not contain a monophyletic ingroup and some had few internal branches with non-zero length, we were able to compute the first two variables for 2,319 trees (a success rate of 54.7%).

To find the sequences best suited for molecular dating, we removed the remaining chunks associated with the shortest 10% as well as the longest 10% of trees. This reduced the range of tree length in terms of expected substitutions per site from (8.48×10^{-7}, 141) to (0.111, 20.9). In the next step, we excluded the least clock-like one-third of all remaining trees, reducing the maximum root-to-tip branch length variance observed in the dataset from 40.2 to 1.78×10^{-3}. Finally, we dropped all chunks whose trees were below the ninetieth percentile of bipartition support, raising the minimum congruence with the reference tree (as measured by the proportion of shared bipartitions) from 0 to 0.0345 and reducing the number of retained chunks to 66. Of these chunks, 65 were 50 bp long and one was 47 bp long.

We concatenated the 66 chunks from the filtered dataset into a single alignment using SequenceMatrix 1.8 (ref. [58]), which we then analysed using PartitionFinder 2.0 (ref. [59]) under the 'greedy' search option and the BEAST model set. The search yielded a final set of 14 partitions whose lengths ranged from 50 to 497 bp. We excluded the two shortest partitions (both 50 bp long) from subsequent analyses to ensure that a sufficient amount of data was available for estimating the parameters of a separate substitution model and estimated divergence times using this SortaDate-filtered alignment following procedures identical to those described above (Supplementary Fig. 8).

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Code availability. PHYLUCE source code is available under an open source license from https://github.com/faircloth-lab/phyluce/, and other custom scripts used to process these data are available under a CC0 license from Dryad (https://doi.org/10.5061/dryad.085dd).

Data availability. Raw sequence read data and contig assemblies are available from NCBI BioProject (PRJNA348720). Assembled contigs are also available from the NCBI Targeted Locus Study database KALI00000000-KAPX00000000. Individual Sequence Read Archive and Targeted Locus Study accession numbers are provided in Supplementary Tables 1 and 4. Sequence alignments, assembled contigs, phylogenetic trees, records of data-processing steps, scripts and related files are available from Dryad (https://doi.org/10.5061/dryad.085dd).

Received: 15 November 2016; Accepted: 6 February 2018;
Published online: 12 March 2018

References

1. Erwin, D. H. Lessons from the past: biotic recoveries from mass extinctions. *Proc. Natl Acad. Sci. USA* **98**, 5399–5403 (2001).
2. Krug, A. Z., Jablonski, D. & Valentine, J. W. Signature of the end-Cretaceous mass extinction in the modern biota. *Science* **323**, 767–771 (2009).
3. Jarvis, E. D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
4. Prum, R. O. et al. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569–573 (2015).
5. O'Leary, M. A. et al. The placental mammal ancestor and the post-K–Pg radiation of placentals. *Science* **339**, 662–667 (2013).
6. Longrich, N. R., Bhullar, B.-A. S. & Gauthier, J. A. Mass extinction of lizards and snakes at the Cretaceous–Paleogene boundary. *Proc. Natl Acad. Sci. USA* **109**, 21396–21401 (2012).
7. Benton, M. J. Diversification and extinction in the history of life. *Science* **268**, 52–58 (1995).
8. Near, T. J. et al. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc. Natl Acad. Sci. USA* **110**, 12738–12743 (2013).
9. Sanciangco, M. D., Carpenter, K. E. & Betancur-R, R. Phylogenetic placement of enigmatic percomorph families (Teleostei: Percomorphaceae). *Mol. Phylogenet. Evol.* **94**, 565–576 (2016).
10. Santini, F., Harmon, L. J., Carnevale, G. & Alfaro, M. E. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC Evol. Biol.* **9**, 194 (2009).
11. Near, T. J. et al. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl Acad. Sci. USA* **109**, 13698–13703 (2012).
12. Patterson, C. & Smith, A. B. Periodicity in extinction: the role of systematics. *Ecology* **70**, 802–811 (1989).
13. MacLeod, N. et al. The Cretaceous–Tertiary biotic transition. *J. Geol. Soc. Lond.* **154**, 265–292 (1997).
14. Cavin, L. in *Geological and Biological Effects of Impact Events* (eds Buffetaut, E. & Koeberl, C.) 141–158 (Springer, Berlin & Heidelberg, 2002).
15. Friedman, M. Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proc. R. Soc. B* **277**, 1675–1683 (2010).
16. Friedman, M. Ecomorphological selectivity among marine teleost fishes during the end-Cretaceous extinction. *Proc. Natl Acad. Sci. USA* **106**, 5218–5223 (2009).
17. Sibert, E. C. & Norris, R. D. New age of fishes initiated by the Cretaceous–Paleogene mass extinction. *Proc. Natl Acad. Sci. USA* **112**, 8537–8542 (2015).
18. Patterson, C. An overview of the early fossil record of the acanthomorphs. *Bull. Mar. Sci.* **52**, 29–59 (1993).
19. Johnson, D. G. Percomorph phylogeny: progress and problems. *Bull. Mar. Sci.* **52**, 3–28 (1993).
20. Johnson, D. G. & Patterson, C. Percomorph phylogeny: a survey of acanthomorphs and a new proposal. *Bull. Mar. Sci.* **52**, 554–626 (1993).
21. Davesne, D. et al. The phylogenetic intrarelationships of spiny-rayed fishes (Acanthomorpha, Teleostei, Actinopterygii): fossil taxa increase the congruence of morphology with molecular data. *Front. Ecol. Evol.* **4**, 129 (2016).
22. Alfaro, M. E. et al. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Natl Acad. Sci. USA* **106**, 13410–13414 (2009).
23. Nelson, J. S. *Fishes of the World* 4th edn (Wiley, Hoboken, 2006).
24. Betancur, R. et al. The tree of life and a new classification of bony fishes. *PLoS Curr.* **5**, https://doi.org/10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288 (2013).
25. Wainwright, P. C. et al. The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. *Syst. Biol.* **61**, 1001–1027 (2012).

26. Faircloth, B. C., Sorenson, L., Santini, F. & Alfaro, M. E. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLoS ONE* **8**, e65923 (2013).
27. Bejerano, G. et al. Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
28. Faircloth, B. C. et al. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* **61**, 717–726 (2012).
29. Setiamarga, D. H. E. et al. Divergence time of the two regional medaka populations in Japan as a new time scale for comparative genomics of vertebrates. *Biol. Lett.* **5**, 812–816 (2009).
30. Yamanoue, Y., Miya, M., Inoue, J. G., Matsuura, K. & Nishida, M. The mitochondrial genome of spotted green pufferfish *Tetraodon nigroviridis* (Teleostei: Tetraodontiformes) and divergence time estimation among model organisms in fishes. *Genes Genet. Syst.* **81**, 29–39 (2006).
31. Yang, Z. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226 (2006).
32. Zheng, Y., Peng, R., Kuro-o, M. & Zeng, X. Exploring patterns and extent of bias in estimating divergence time from mitochondrial DNA sequence data in a particular lineage: a case study of salamanders (order Caudata). *Mol. Biol. Evol.* **28**, 2521–2535 (2011).
33. Brown, J. W. & Van Tuinen, M. (eds) in *Living Dinosaurs: The Evolutionary History of Modern Birds* 306–324 (Wiley, Chichester, 2011).
34. Smith, B. T. & Klicka, J. Examining the role of effective population size on mitochondrial and multilocus divergence time discordance in a songbird. *PLoS ONE* **8**, e55161 (2013).
35. Friedman, M. & Sallan, L. C. Five hundred million years of extinction and recovery: a Phanerozoic survey of large-scale diversity patterns in fishes. *Palaeontology* **55**, 707–742 (2012).
36. Miya, M. et al. Evolutionary origin of the scombridae (tunas and mackerels): members of a Paleogene adaptive radiation with 14 other pelagic fish families. *PLoS ONE* **8**, e73535 (2013).
37. Harrington, R. C. et al. Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. *BMC Evol. Biol.* **16**, 224 (2016).
38. Carnevale, G. & Johnson, G. D. A Cretaceous cusk-eel (Teleostei, Ophidiiformes) from Italy and the Mesozoic diversification of percomorph fishes. *Copeia* **103**, 771–791 (2015).
39. Schwarzhans, W. in *Mesozoic Fishes—Systematics and Paleoecology* (ed. Viohl, G.) 417–431 (F. Pfeil, München, 1996).
40. Faircloth, B. C. & Glenn, T. C. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS ONE* **7**, e42543 (2012).
41. Glenn, T. C. et al. Adapterama I: universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). Preprint at https://doi.org/10.1101/049114 (2016).
42. Blumenstiel, B. et al. Targeted exon sequencing by in-solution hybrid selection. *Curr. Protoc. Hum. Genet.* **66**, 18.4.1–18.4.24 (2010).
43. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
44. Faircloth, B. C. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* **32**, 786–788 (2015).
45. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701 (2012).
46. Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol. Biol.* **14**, 82 (2014).
47. Aberer, A. J., Kobert, K. & Stamatakis, A. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Mol. Biol. Evol.* **31**, 2553–2556 (2014).
48. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
49. Phillips, M. J. & Penny, D. The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylogenet. Evol.* **28**, 171–185 (2003).
50. Phillips, M. J., Delsuc, F. & Penny, D. Genome-scale phylogeny and the detection of systematic biases. *Mol. Biol. Evol.* **21**, 1455–1458 (2004).
51. Seo, T.-K. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Mol. Biol. Evol.* **25**, 960–971 (2008).
52. Mirarab, S. & Warnow, T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, i44–i52 (2015).
53. Sukumaran, J. & Holder, M. T. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26**, 1569–1571 (2010).
54. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
55. Dos Reis, M. & Yang, Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol. Biol. Evol.* **28**, 2161–2172 (2011).
56. Smith, S., Brown, J. W. & Walker, J. F. So many genes, so little time: comments on divergence-time estimation in the genomic era. Preprint at https://doi.org/10.1101/114975 (2017).
57. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
58. Vaidya, G., Lohman, D. J. & Meier, R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* **27**, 171–180 (2011).
59. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017).
60. Chen, W.-J. et al. New insights on early evolution of spiny-rayed fishes (Teleostei: Acanthomorpha). *Front. Mar. Sci.* **1**, 1–17 (2014).

Acknowledgements

We thank J. Johnson for the illustrations used in the figures. This work was partially supported by National Science Foundation grants DEB-0842397 (to M.E.A.); DEB-1242260, DEB-1655624 (to B.C.F.); and startup funds from Louisiana State University (to B.C.F.). DNA alignment was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource Center—a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology. Phylogenetic analysis portions of this research were conducted using high-performance computing resources provided by Louisiana State University (http://www.hpc.lsu.edu).

Author contributions

M.E.A. conceived the study. M.E.A., B.C.F., R.C.H., M.F. and T.J.N. designed the study. L.S., R.C.H. and B.C.F. collected the sequence data from enriched UCE loci. M.F. selected appropriate calibration points. B.C.F., M.E.A., D.Č. and C.H.O. performed the data analyses. M.E.A. and B.C.F. wrote the manuscript with help from R.C.H., L.S., M.F., D.Č. and T.J.N. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0494-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.E.A. or B.C.F.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Michael Alfaro, Brant Faircloth

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

Describe how sample size was determined.

At the scale of this phylogenetic analysis a single individual per species is sufficient to draw inferences about the evolutionary history of major acanthomorph lineages.

2. Data exclusions

Describe any data exclusions.

No data were excluded from this study.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

We report all protocols for the generation of phylogenomic data and have deposited data and pipeline descriptions for others to reproduce our contig assembly, alignment, and downstream phylogenetic analyses steps.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

We performed a historical evolutionary analyses that did not allow for randomization.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

We performed a historical evolutionary analysis and so blinding was not relevant.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☒	☐	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	A statement indicating how many times each experiment was replicated
☒	☐	The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐	Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	We used the Phyluce pipeline as well as the phylogenetics programs RAXML, EXABAYES, ASTRAL, and PAML/MCMCTree for evolutionary analyses. Citations to all software used are given in the manuscript, and scripts to replicate the analysis along with input files have been deposited in Dryad.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.	All data and results from this study are available to the community through NCBI SRA, NCBI Genbank, and Dryad.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No anitbodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No cell lines were used.
b. Describe the method of cell line authentication used.	No cell lines were used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No cell lines were used.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.	DNA samples of animals used in this study were obtained from natural history museums or collected following IACUC approval (UCLA IACUC #2008-176-21).

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	This study did not involve human research participants.