nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-018-0719-8

OPEN

The genome of the jellyfish *Aurelia* and the evolution of animal complexity

David A. Gold[1,2,12*], Takeo Katsuki[3,12*], Yang Li[4], Xifeng Yan[4], Michael Regulski[5], David Ibberson[6], Thomas Holstein[7], Robert E. Steele[8], David K. Jacobs[9] and Ralph J. Greenspan[3,10,11,12*]

We present the genome of the moon jellyfish *Aurelia*, a genome from a cnidarian with a medusa life stage. Our analyses suggest that gene gain and loss in *Aurelia* is comparable to what has been found in its morphologically simpler relatives—the anthozoan corals and sea anemones. RNA sequencing analysis does not support the hypothesis that taxonomically restricted (orphan) genes play an oversized role in the development of the medusa stage. Instead, genes broadly conserved across animals and eukaryotes play comparable roles throughout the life cycle. All life stages of *Aurelia* are significantly enriched in the expression of genes that are hypothesized to interact in protein networks found in bilaterian animals. Collectively, our results suggest that increased life cycle complexity in *Aurelia* does not correlate with an increased number of genes. This leads to two possible evolutionary scenarios: either medusozoans evolved their complex medusa life stage (with concomitant shifts into new ecological niches) primarily by re-working genetic pathways already present in the last common ancestor of cnidarians, or the earliest cnidarians had a medusa life stage, which was subsequently lost in the anthozoans. While we favour the earlier hypothesis, the latter is consistent with growing evidence that many of the earliest animals were more physically complex than previously hypothesized.

A goal of comparative genomics is to decipher the causal connections between genome composition and animal form. The phylum Cnidaria (sea anemones, corals, hydroids and jellyfish) holds a pivotal place in such studies. Phylogenetic analyses consistently support cnidarians as the sister clade to Bilateria (protostomes plus deuterostomes), the clade that encompasses 99% of extant animals (Fig. 1a)[1,2]. Putative fossils of extant cnidarian classes have been identified in lower Cambrian strata, suggesting that cnidarian diversification represents one of the oldest evolutionary events among living animal phyla[3,4]. Nearly all cnidarian life cycles incorporate polyp and/or medusa body plans (Fig. 1b), the former a sessile life stage, and the latter a swimming predator equipped with neural and sensory structures that rival those of many bilaterians. Sequenced cnidarian genomes include the sea anemones *Nematostella vectensis*[5] and *Exaiptasia pallida* (syn. *Aiptasia* sp.)[6], the coral *Acropora digitifera*[7] and the hydroid *Hydra vulgaris* (formerly *Hydra magnipapillata*)[8]. However, none of these species has a medusa life stage, and thus a major event in the evolution of complex animal life has not been subjected to whole genome sequencing.

To improve our understanding of life history evolution in cnidarians, we have generated a draft genome assembly from the moon jellyfish *Aurelia* ('species 1' strain sensu, Dawson and Jacobs[9]), augmented with transcriptomes that cover the major life stages. *Aurelia* offers a tractable laboratory model and a valuable addition to comparative genomics. It is a member of the medusozoan class Scyphozoa, which represents a sister clade to *Hydra* and its relatives (Hydrozoa)[10]. The *Aurelia* medusa is a swimming planktivore, featuring complex neural and sensory system architecture manifested in eight structures called rhopalia, which are located on the margin of the medusa's bell (Fig. 1c,d). The rhopalium features multiple sensory structures—including an eye-cup, a mechanosensory touch plate and a geosensory statocyst—and is patterned using several genes involved in bilaterian sensory organogenesis[11,12]. No comparable sensory structures exist in *Nematostella*, *Exaiptasia*, *Acropora* or *Hydra*. Genomes from medusa-bearing cnidarians such as *Aurelia*—alongside the forthcoming *Clytia* genome[13]—thus provide a new vantage into the evolution of complex animal life cycles.

Results and discussion

We sequenced and assembled the *Aurelia* genome using a combination of Illumina paired-end, mate-pair and PacBio data (see Methods section). Our final assembly has a total size of 713 megabases (Mb), which is consistent with previous estimates of the size of the *Aurelia* genome (C-value = 0.73 pg)[14]. This makes the *Aurelia* genome larger than sequenced anthozoan genomes, but smaller than some strains of *H. vulgaris* (~1.1–1.35 Gb for brown hydra and ~0.38 Gb for green hydra; see Supplementary Table 1)[5–8]. The *Aurelia* assembly is more fragmented than the anthozoan genomes. This is largely due to a high percentage of repetitive DNA, with transposable elements making up ~49.5% of the genome, and another ~0.8% of the genome consisting of simple tandem repeats (see Supplementary Table 5 for a summary of transposable elements). Synteny analysis performed with MCScanX[15] suggests that antho-

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [2]Department of Earth and Planetary Sciences, University of California Davis, Davis, CA, USA. [3]Kavli Institute for Brain and Mind, University of California San Diego, La Jolla, CA, USA. [4]Department of Computer Science, University of California Santa Barbara, Santa Barbara, CA, USA. [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [6]Deep Sequencing Core Facility, Cell Networks, Heidelberg University, Heidelberg, Germany. [7]Department of Molecular Evolution and Genomics, Centre for Organismal Studies, Heidelberg University, Heidelberg, Germany. [8]Department of Biological Chemistry and Developmental Biology Center, University of California Irvine, Irvine, CA, USA. [9]Department of Ecology and Evolution, University of California Los Angeles, Los Angeles, CA, USA. [10]Division of Biological Sciences, University of California San Diego, La Jolla, CA, USA. [11]Department of Cognitive Science, University of California San Diego, La Jolla, CA, USA. [12]These authors contributed equally: D. A. Gold, T. Katsuki, R. J. Greenspan. *e-mail: dgold@ucdavis.edu; takeo.katsuki@gmail.com; rgreenspan@ucsd.edu

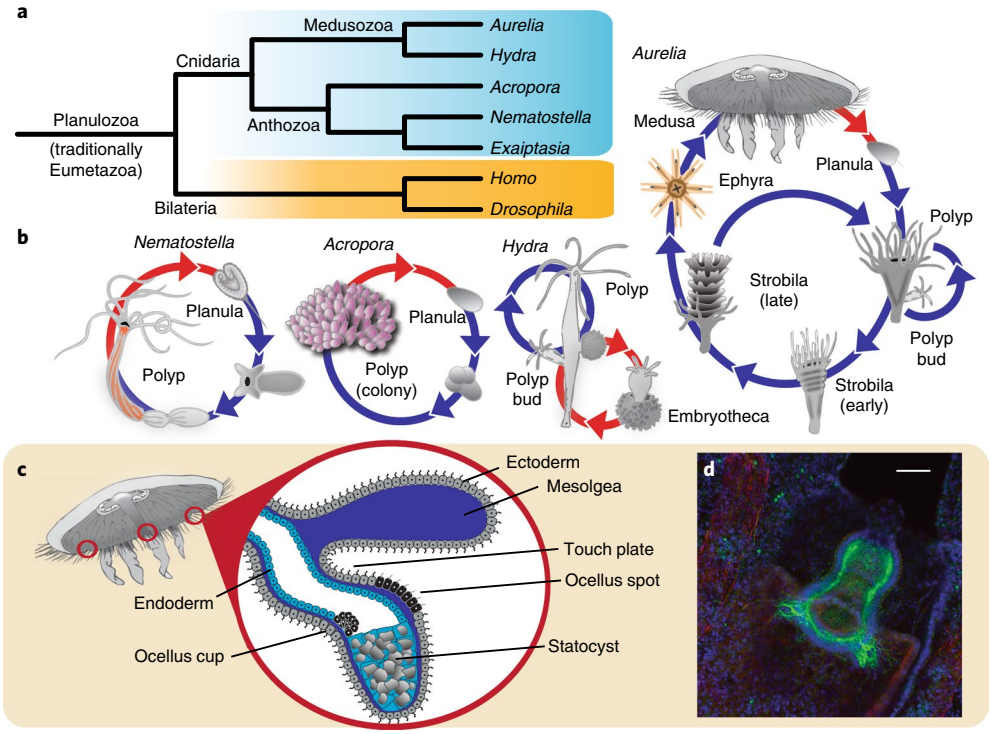

Fig. 1 | Cnidarian relationships, life cycles and sensory structures. a, Cladogram showing the phylogenic relationships between cnidarians with published genome sequences. **b**, Representative life cycles for cnidarians. Red arrows indicate sexual reproduction; blue arrows indicate metamorphosis and/or asexual reproduction. Some images modelled after Technau and Steele[87]. **c**, Organization of the rhopalia, a sensory structure found only in certain medusozoans such as *Aurelia*. **d**, Antibody staining demonstrating the clustering of tyrosinated tubulin-positive neurons (green) in the rhopalia. Red, phalloidin (actin stain); green, tyrosinated tubulin (Sigma, cat. no. T9028); blue, TO-PRO-3 Iodide (nuclear stain). Scale bar, 50 µm.

zoans share far more syntenic blocks of orthologous genes amongst themselves than they do with *Aurelia* (see Supplementary Table 6 and the Supplementary Data). However, *Aurelia* shares more syntenic gene blocks with anthozoans than it does with *Hydra*, which suggests that its genome architecture is less derived. We found no evidence for trans-spliced leader sequences in our messenger RNA models, meaning that their presence in some hydrozoans is probably a clade-specific novelty[16,17]. Overall, the *Aurelia* genome shares characteristics with both anthozoans and hydrozoans, consistent with its phylogenetic placement (Fig. 1a).

Our annotation pipeline resulted in 29,964 gene models. This is on the higher end of gene count estimates in early branching animals, but is fewer than recent estimates for *Acropora* (Supplementary Table 1) and far fewer than the >40,000 genes currently predicted in the sponge *Amphimedon*[18,19]. Benchmarking Universal Single-Copy Ortholog (BUSCO)[20] analysis of these gene models recovers complete or partial sequences for 76% of 'core' metazoan genes and 86% of 'core' eukaryotic genes, making the *Aurelia* assembly comparable to early branching organisms such as *Amphimedon*, *Nematostella* and *Mnemiopsis* (see Extended Data Table 3 in Levin et. al[21], and the Supplementary Data for detailed BUSCO output). Using Pfam annotation, we catalogued the number of proteins with putative transcription-factor and peptide-signalling domains (Supplementary Tables 8 and 9; see the Supplementary Data for full Pfam annotation). In nearly every case, the numbers of conserved proteins in *Aurelia* fall within the range of other cnidarians. Based on these results, we feel confident that we have generated a draft genome of sufficient quality for comparative study.

The first question we wanted to address was intraspecies variability across *Aurelia* populations. The jellyfish used in our research, which is native to the coastline of California, is commonly referred to as *Aurelia aurita*. However, genetic markers reveal large sequence differences between various *Aurelia* populations (up to 40% divergence in ITS-1 and 23% in cytochrome c oxidase subunit I (CO1))[9]. Such diversity is comparable to interspecific differences in other marine animals, and suggests that the *Aurelia* species complex is ancient, probably originating in the Mesozoic[9,22]. Do these large differences in mitochondrial and non-coding regions imply equally large changes at the peptide level? To test this, we compared the protein models from our Californian strain of *Aurelia* to previously published transcriptomes from populations in Roscoff, France[23], and Eilat, Israel[24]. The complete mitochondrial genome of our organism (contig 'Seg3751') shows 99% similarity to the '*Aurelia aurita* (2)' mitogenome published by Park et al. (National Center for Biotechnology Information (NCBI) accession HQ694729)[22]. Phylogenetic analysis of the CO1 sequence derived from this mitogenome confirms that our strain is part of the 'species 1' complex (Fig. 2a). CO1 sequences of the Californian and Roscoff strains are ~97.8% identical, while the Californian and Eilat strains are ~81.5% identical. The average pair-wise identity between single-copy orthologous proteins is consistent with the CO1 results; amino acid sequences from the California and Roscoff strains are, on average, ~97.7% identical, while the California and Eilat strains are ~90.9% identical (Fig. 2b). For comparison, these same proteins in mice (*Mus musculus*) and rats (*Rattus norvegicus*) are, on average, ~95.1% identical (see the Supplementary Data). This means there is greater protein sequence divergence between some *Aurelia* populations than there is between mice and rats. These results suggest that, similar to *Hydra*, substantial variation exists across *Aurelia* genomes.

As the first step in our comparison of the *Aurelia* genome to other cnidarian genomes, we used OrthoFinder[25] to group the cnidarian proteomes—as well as the bilaterians *Branchiostoma*, *Capitella*, *Drosophila*, *Homo*, *Lottia* and *Limulus*—into putative sets

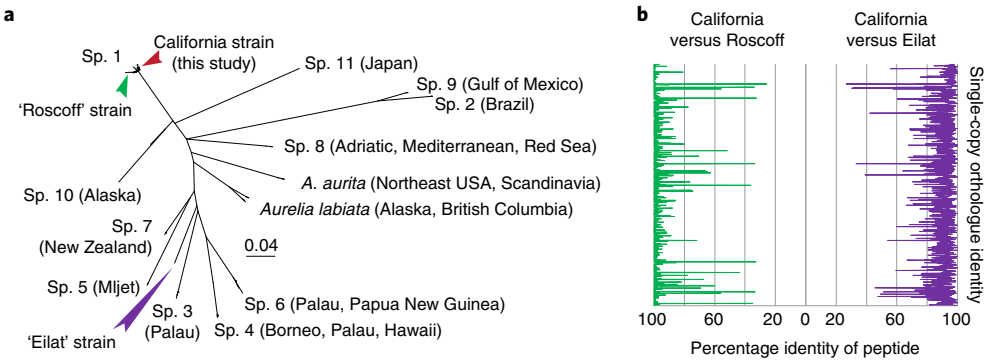

Fig. 2 | Intraspecies variability across the genus *Aurelia*. a, Unrooted phylogenetic tree of *Aurelia* strains based on the CO1 genetic marker. Our 'California stain' is noted with a red arrow; the 'Roscoff' and 'Eilat' strains are noted with green and purple arrows, respectively. **b**, A graph showing the percentage amino acid identity of peptides between the strains of *Aurelia*. This analysis is restricted to single-copy orthologues shared between the three strains.

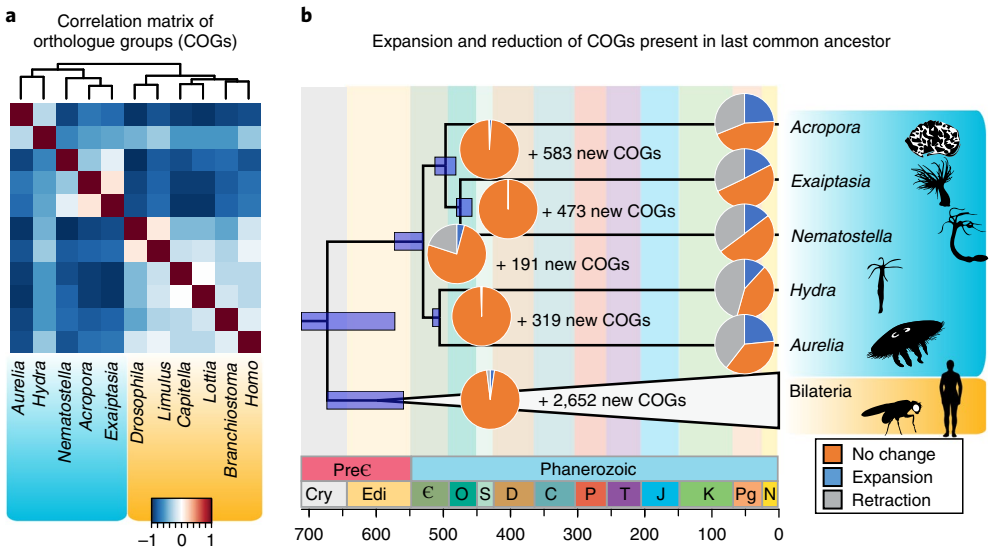

Fig. 3 | Gene expansions and losses among sequenced cnidarian genomes. a, A correlation matrix of orthologous gene clusters, represented as heat maps. The heat map codes clusters as binary 'present/absent' data for each taxon. **b**, A molecular clock for the five cnidarians and six bilaterians included in this study. Pie charts represent 8,263 conserved gene families present in the last common ancestor of cnidarians and bilaterians; the colours in each chart represent the number of families experiencing gene copy expansion, retraction or no change at that evolutionary node. Nodes within the Bilateria have been removed for simplicity, but all data for node dates and expansion/contraction statistics are available in the Supplementary Data. Abbreviations for the *x* axis: Cry, Cryogenian; Edi, Ediacaran; Є, Cambrian; O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene; Ng, Neogene.

of conserved orthologues. *Aurelia* shares 378 conserved orthologous groups (COGs) with 1 or more bilaterians to the exclusion of other cnidarian genomes, including 27 COGs shared with *Drosophila* and 60 COGs with humans (Supplementary Fig. 2; the full list is provided in the Supplementary Data). Noteworthy, vetted members of this list include homologues of FBXO25/FBXO32 and RAG1—members of the FoxO signalling pathway that regulates stem cell maintenance in *Hydra*[26,27]—as well as JMY, which dynamically regulates cell motility and P53-based tumour suppression[28]. RAG1 has previously been identified in the hydrozoan jellyfish *Podocoryna*[29], which suggests that the FoxO pathway might be broadly conserved across medusa-bearing cnidarians. Despite the hypothesized derived nature of medusozoans, their orthologue repertoire is equally similar to bilaterians compared to anthozoans (Fig. 3a); this suggests that medusozoans and anthozoans have retained comparable portions of the ancestral cnidarian/bilaterian gene repertoire.

Focusing on orthologue clusters shared between cnidarians and bilaterians, we next traced patterns of gene gain and loss across 8,263 conserved gene families shared in the cnidarian/bilaterian (planulozoan) last common ancestor (Fig. 3b). Our results suggest that cnidarians and bilaterians each had their own pattern of gene expansions and contractions, as well as lineage-specific increases in novel gene families. This is consistent with the correlation matrix (Fig. 3a), which suggests that the organisms in our data set have largely dissimilar patterns of gene gain and loss compared with each other. The fraction of gene family contractions in *Aurelia* inherited from the planulozoan last common ancestor (~40%) is slightly higher than anthozoans (31–35%) but lower than *Hydra* (46%), which has undergone substantial gene loss. Regarding gene expansions, the

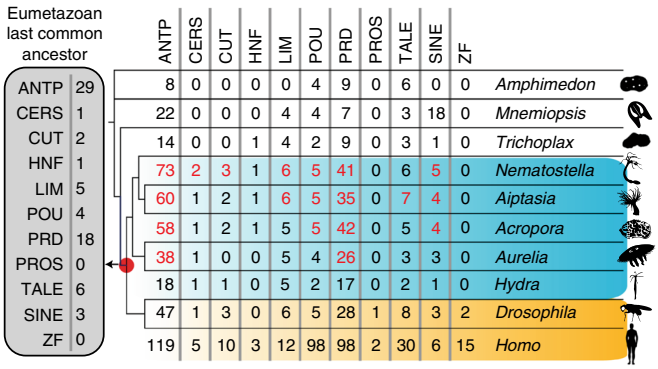

Fig. 4 | The homeodomain complement of various animals, divided into the 11 major classes proposed by Zhong and Holland[88]. Rows represent candidate genomes from major animal groups, organized by their evolutionary relationships. Columns contain gene counts for each of the 11 major homeodomain classes. The hypothesized complement of the cnidarian/bilaterian last common ancestor is presented in the grey box to the left. Increases in cnidarian gene counts are noted in red. Gene counts for non-cnidarians are taken from HomeoDB2[88] and refs [30,31,89].

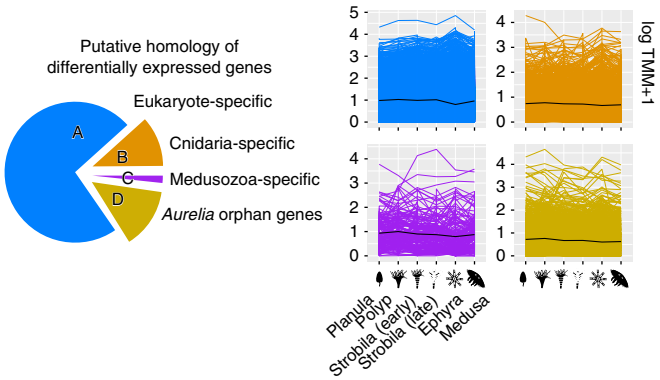

Fig. 5 | RNA-seq expression profiles across the life cycle. Breakdown of 11,963 differentially expressed genes across the *Aurelia* life cycle by their putative taxonomic origin (left), and by their associated gene expression profiles (right). The gene expression profiles are organized by life stage on the *x* axis. The *y* axis shows the log transcript per million (TMM) counts for each gene in the cluster.

rate in *Aurelia* (~23%) is comparable to that of available cnidarian genomes (~12–24%). If we expand our consideration to genes not present in the last common ancestor, gene innovation appears to be commonplace in the anthozoans; the number of COGs restricted to 2 or more anthozoans (1,695 clusters) is far greater than the numbers restricted to medusozoans (319 clusters; see Supplementary Fig. 2 for details). There are several sets of transcription factors that appear greatly expanded in *Aurelia* compared with other cnidarians, including proteins featuring a basic region leucine zipper, C2H2 type zinc finger, ETS, GATA zinc finger and/or HMG box domain (Supplementary Table 8). In all of these cases, many of the genes are differentially expressed, and demonstrate complex expression profiles across *Aurelia's* life history (Supplementary Figs. 3 and 4). These gene expansions provide possible candidates for regulating the complex life cycle found in *Aurelia*, and are worthy of future study. But at a genome-wide vantage, there is little evidence that the expansion of conserved genes played an outsized role in the evolution of medusozoan body-plans.

Homeobox genes—a large clade of transcription factors that share a ~60-peptide DNA-binding homeodomain region—are primary candidates in the study of animal body-plan evolution, and a common starting point when analysing the gene content of early branching animal lineages[30–33]. In our list of COGs, we recovered several homeobox genes that *Aurelia* putatively shares with bilaterians to the exclusion of available cnidarian genomes. However, high sequence conservation within this gene group limits vetting with the Basic Local Alignment Search Tool (reciprocal-BLAST), so we performed a more detailed analysis of homeobox evolution using phylogenetic analysis (see Methods section). We attribute cnidarian homeodomains to 69 bilaterian families encompassing 9 classes (Fig. 4), which significantly increases the reconstructed homeobox gene complement of the planulozoan last common ancestor[32]. Anthozoans have higher homeobox gene counts than medusozoans; this is partly attributable to gene loss in medusozoans, but is mostly the result of multiple rounds of anthozoan-specific gene duplication events[32,34]. Putative anthozoan expansions involve Dmbx-, POU3-, Barx-, Bari-, Nk2- and Noto-like genes, as well as large radiations of PRD- and ANTP-class genes that cannot be readily matched to bilaterian genes (Supplementary Table 10 and see the Supplementary Data for homeodomain trees and assignments). In contrast, *Aurelia* appears to be missing 21 homeodomains found in 1 or more anthozoans (17 of which are also missing in *Hydra*), while it had mild expansions of Otx-, Vsx- and Hox9-13/15-like genes. These results provide a case study where the anthozoan gene repertoire is larger than that of *Aurelia*, despite the latter's complex life cycle.

Given that conserved gene families are not broadly expanded in *Aurelia*, it is nevertheless possible that taxonomically restricted (orphan) genes have played a driving role in the evolution of medusozoan life stages. To test this hypothesis, we analysed RNA sequencing (RNA-seq) data from six stages in the *Aurelia* life cycle: planula, polyp, early strobila, late strobila, ephyra and juvenile medusa (Fig. 1a). A total of 11,963 differentially expressed genes were phylogenetically annotated based on a series of BLAST queries (results provided in the Supplementary Data). We found no evidence that taxonomically restricted genes demonstrate a collective trend towards upregulation in taxonomically restricted life stages (Fig. 5). Instead, genes unique to *Aurelia* are expressed more or less evenly across the life cycle. Some orphan genes are likely to play important roles in the development of the medsua[23] but, at a transcriptome-wide level, the evolution of novel life stages in *Aurelia* appears to be the result of redeploying deeply conserved genes as opposed to acquiring new ones.

Since it appears that the development of medusozoan life stages involves redeployment of conserved genes, we next asked whether these genes demonstrate evidence of conserved functionality. We first searched for transcripts that are differentially regulated between pan-cnidarian life stages (planula through polyp) and medusozoan-specific life stages (early strobila through medusa). This analysis was restricted to genes that were successfully annotated using the Uniprot Swissprot[35] data set. Enriched gene ontology annotations from these two clusters (provided in the Supplementary Data) are consistent with recent research on *Aurelia* development; for example, that the polyp-to-medusa transition involves major changes in the nervous system[36], musculature[37] and cnidocyte composition[38]. In a separate analysis, we annotated these differentially expressed genes based on their best BLAST hits from the *Drosophila* or *Homo* proteomes (see the Supplementary Data). These annotated genes were clustered into expression profiles (Supplementary Fig. 9) and submitted to STRING v10[39] to look for the possible conservation of protein–protein interactions and enriched gene networks. According to STRING, all clusters contain significantly more protein–protein interactions than expected by chance (protein–protein interaction enrichment *P* value >0.05). These results support the hypothesis that conserved, differentially expressed genes in the

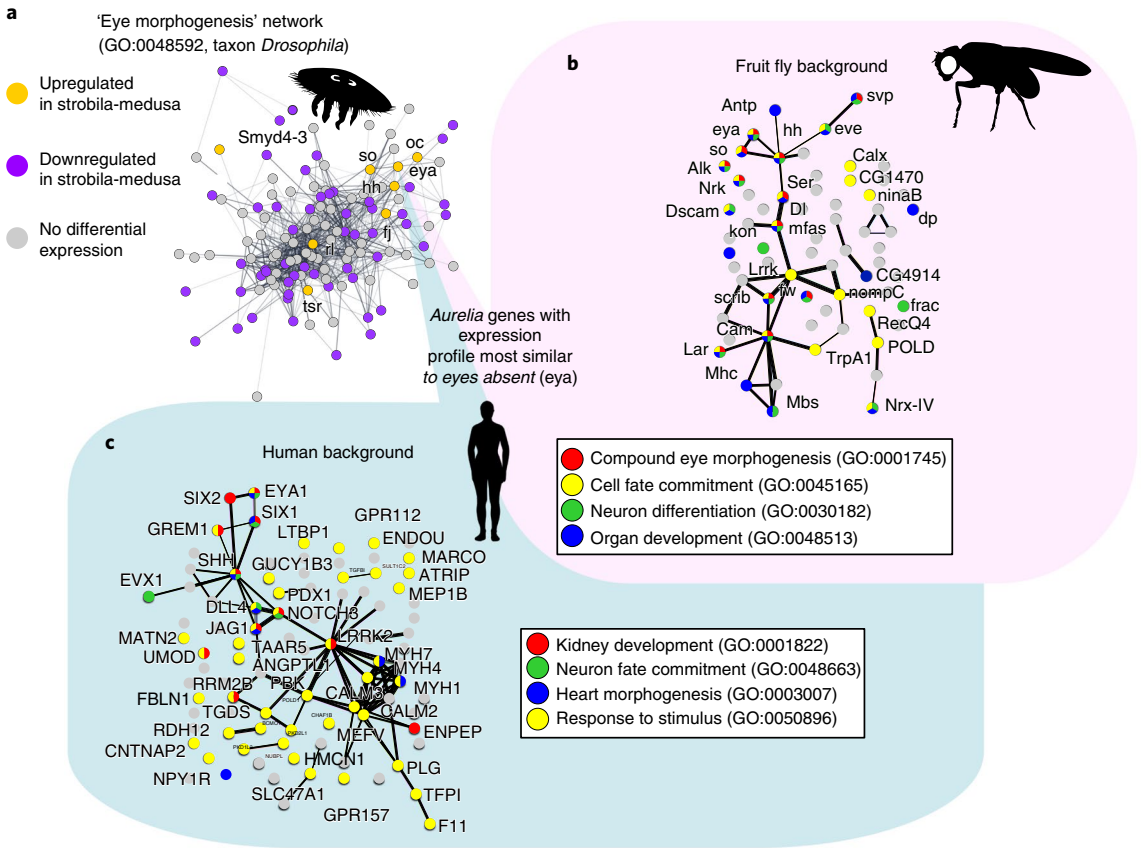

Fig. 6 | Clustering of differentially expressed genes and gene ontology (GO) terms. a, A protein interaction network showing genes involved in *Drosophila* eye development. The circles are coloured based on their expression profile in *Aurelia*. **b**,**c**, Protein interaction networks and select enriched GO terms for the *Aurelia* genes most similar in expression profile to *eyes absent*. Networks and enrichment analysis were performed using STRING, and based on putative homology to proteins in *Drosophila* (**b**) and humans (**c**). The illustrated GO terms were chosen by how informative they are and their non-redundancy. A full list of proteins and enriched GO terms are provided in the Supplementary Data.

medusa life stages are frequently involved in gene networks present in bilaterian animals.

For a final analysis, we focused on the enrichment of eye development proteins, because the homology between bilaterian and cnidarian eyes has been the subject of a long-standing debate in evolutionary biology[40]. *Aurelia* rhopalia feature a simple 'pit eye' that is probably capable of recognizing the direction of light[41] (Fig. 1c), and scyphozoans are the sister taxon to cubozoans (box jellies), which feature complex eyes with a lens and retina. We began our analysis by using QuickGO to collect all *Drosophila* proteins known to play a role in eye morphogenesis (see the Supplementary Data). We created an interaction network for these proteins using STRING, and coloured them based on their expression profile in *Aurelia* (Fig. 6a). Of the genes involved in *Drosophila* eye morphogenesis, 61% have a homologue in *Aurelia* (292/478 queries); of these, ~59% exhibit significant differential expression in *Aurelia* (172/292 queries). For the 172 differentially expressed genes, only 19 are upregulated in medusozoan-specific life stages. These results suggest that proteins involved in *Drosophila* eye morphogenesis are not uniformly upregulated in *Aurelia*, and that many aspects of eye development are unlikely to be conserved.

Despite the abovementioned results, many of the major players of the 'canonical' eye-patterning network are upregulated in *Aurelia* during development of the medusa, including *sine oculis* (so), *eyes absent* (eya) and *ocelliless* (oc) (Fig. 6a). Many of these genes have previously been shown to be expressed in the *Aurelia* rhopalia[11,42]. We therefore flipped our original question; instead of asking what bilaterian eye-patterning genes are conserved in *Aurelia*, we asked, what are the functions of putative *Aurelia* eye-patterning genes in bilaterians? We used our gene clustering analysis to extract the genes with most similar expression profiles to *eyes absent* (Supplementary Fig. 8). Based on putative homologues in *Drosophila* and humans, we looked for potential conserved protein interactions and enriched gene ontologies. When compared against the *Drosophila* proteome, the *Aurelia* genes with expression profiles most similar to *eyes absent* are enriched in functions involving neurogenesis and compound eye formation (Fig. 6b). This analysis revealed some candidate genes for eye development in *Aurelia* that were missed in the QuickGO analysis. Interestingly, the same set of genes does not show enrichment for eye development in humans; instead, the list is dominated by proteins involved in kidney/nephron formation, neuron commitment and heart morphogenesis (Fig. 6c). Overall, our results provide intriguing evidence that sensory structures in *Aurelia* share 'deep homology' with bilaterian organs via ancestral multifunctional cell types[43,44], and provide a case study for how the *Aurelia* genome can be queried to study gene regulatory network evolution in animals.

Conclusion

In conclusion, our results do not support the hypothesis that an increase in life history complexity in cnidarians is associated with an increase in gene number. Instead, *Aurelia* appears to pattern its strobila, ephyra and medusa life stages using many of the same genes found in bilaterian animals, possibly through the redeploy-

ment and modification of ancestral gene networks. This finding adds to a growing body of evidence that the evolution of the medusa life stage required the co-option of previously existing developmental gene networks and cell types. For example, Kraus and colleagues examined the expression of ten pan-metazoan genes in *Aurelia*, and determined that the medusa's bell demonstrates a similar expression profile to the polyp tentacle[45]. The fact that a similar pattern is observed in the hydrozoan *Clytia* led these authors to conclude that medusas are homologous across the Cnidaria, and were derived from the polyp's tentacle analgen[45]. Polyps and medusas of the hydrozoan *Podocoryna* share similar Wnt3/frizzled dynamics, suggesting that axial patterning in the medusa is derived from the polyp[46]. Other structures in the medusa could have even older origins; the eyes of *Cladonema* and *Aurelia* medusae express canonical photoproteins and transcription factors found in bilaterian eyes, suggesting that both may be derived from ancestral photosensitive cells[42,47–49], and light-induced spawning in *Clytia* medusae is driven by a hormone-regulating opsin, which could suggest a deep homology between cnidarian gonadal photosensitive-neurosecretory cells and bilaterian deep brain photoreceptors[50]. While compelling, these studies focus on well-understood and broadly conserved developmental genes, and their results might subsequently overemphasize the similarities between medusae development and the development of other animals. A major contribution of this study to this literature is to demonstrate that these previous observations made on small numbers of genes appear to hold true at a genome-wide vantage.

A second contribution of this study is that it provides the first direct comparison between anthozoan genomes and the genome of a medusa-bearing cnidarian, which led to our discovery that patterns of gene gain, loss and co-option are comparable between the lineages. As important as gene co-option appears in *Aurelia's* evolution, we did discover multiple gene family expansions that could be candidate drivers of medusa development, as well as many taxonomically restricted genes that are upregulated in the polyp-to-medusa transition. This finding is consistent with previous studies that have leveraged high-throughput sequencing to holistically examine medusa development, and broadly support the hypothesis that this life stage is generated from a combination of modified gene regulation as well as gene gain and loss[23,51–53]. However, our analyses allow us to further hypothesize that taxonomically restricted genes are not overrepresented in the polyp-to-medusa transition, and that changes in gene content appear just as common in the anthozoans as they are in *Aurelia*. Although anthozoans such as *Nematostella* are sometimes described as 'basal' cnidarians, this study provides a powerful reminder that all living animals exhibit a mosaic of ancestral and derived traits, and that reconstructing the genomic evolutionary history of animal life will continue to require a broad, comparative approach[54].

We see two ways to interpret our analysis of the *Aurelia* genome, both of which have strong implications for the early evolution of animal life. The first interpretation is that medusozoans evolved a complex life cycle primarily by redeploying genetic and developmental pathways present in the planulozoan last common ancestor. This interpretation, if correct, suggests that animals can transition into radically different ecological niches (in this case, transitioning from benthic to pelagic carnivores) without major innovations in gene content. As the Precambrian–Cambrian transition represents an ecological explosion as much as a morphological one[55], our results challenge the importance of genetic innovations in the early expansion of animal niches. The second possibility is that the last common ancestor of cnidarians had a medusa life stage, which was subsequently lost in anthozoans. This scenario was supported by many studies done in the twentieth century[56], but lost popularity after genetic analyses refuted the hypothesis that hydrozoans are the earliest branching cnidarian lineage. Later cladistic analyses of morphological characters[57] and the derived structure of many medusozoan mitochondrial genomes[58] have been used as additional evidence that the medusozoan body-plan is derived in Cnidaria. However, our results do not support this hypothesis at the genetic level. Despite the current popularity of the 'polyp-first' scenario, it is worth reiterating that neither the polyp nor medusa life stage is found outside of cnidarians; it is therefore equally parsimonious for the first cnidarians to have had a biphasic life cycle that was lost in anthozoans, or for the medusa phase to have originated in medusozoans (see Fig. 1a). Our results cannot distinguish between these two scenarios, but they are consistent with a growing body of literature that the earliest branching animals may have included pelagic carnivores with complex neural and muscular architecture[59,60]. The ecological roles that animals such as jellyfish and ctenophores could have played in Precambrian oceans—where their modern mesoplankton prey were probably absent—is thus a pressing question in studies of the early evolution of animals[61].

In addition to questions of evolution, we anticipate the *Aurelia* genome proving valuable in many other areas of biology. Given the varying degrees of nervous system complexity and behaviour across its life stages, *Aurelia* has and will continue to be an important model for studying the development and function of nervous systems[12]. *Aurelia* is a promising candidate for marine population genomics, as the division of this circumglobal genus into multiple species or subspecies remains unresolved[9]. It is also an important ecological model system, as *Aurelia* is a major culprit in environmentally and economically damaging jellyfish blooms, which may or may not be on the rise due to climate change[62]. Finally, *Aurelia* will provide an important study system in animal regeneration, as different life stages exhibit varying strategies of wound healing[63]. We look forward to additional progress in these fields now that the moon jellyfish has joined the genome family.

Methods

DNA collection and genome assembly. For genome sequencing, a single *Aurelia* polyp obtained from the Birch Aquarium (San Diego, California) was grown into a clonal population in the laboratory. A segment of the mitochondrial CO1 gene was amplified and sequenced, identifying the strain as *Aurelia* sp.1[9]. Polyps were kept in artificial seawater (ASW) at room temperature and fed with *Artemia* nauplii (Brine Shrimp Direct, UT) once every 2 d. Strobilation was induced with 5 µM 5-methoxy-2-methylindole in ASW, or by lowering the temperature of the ASW to 14 °C for about a month. Total DNA was extracted from ephyrae using a salting-out protocol described in the Supplementary Methods. Ephyra were chosen as the source material for genomic DNA collection since multiple ephyra are produced by each polyp, and as pelagic organisms there is a substantially lower risk of collecting the algal contaminants that often grow alongside polyp communities. DNA was sheared to an average size of 10 kbp using a Covaris G-tube. The libraries used and statistics on the sequences obtained are described in the Supplementary Methods and summarized in Supplementary Table 2.

Genome assembly. The strategy for assembling the *Aurelia* genome is illustrated in Supplementary Fig. 1. The 250-bp paired-end reads were assembled into contigs using DISCOVAR de novo with its default options (version 53488, Broad Institute). Only contigs >1 kbp were used for the subsequent scaffolding steps. Initial scaffolding was performed using error-corrected PacBio reads (produced in 2012 using XL-P2 sequencing chemistry) and SSPACE-LR with its default options (version 1-1)[64]. The hybrid error correction of PacBio reads was performed using proovread (version 2.13.8)[65], with error correction based on a combination of 250-bp paired-end reads merged with FLASh[66], as well as high-confidence unitigs generated with ALLPATHS-LG (version 48257)[67]. Unitigs were generated from the 250-bp paired-end reads as a fragment library and the two mate-pair data sets as jumping libraries without quality trimming. ALLPATHS-LG was run with FRAG_COVERAGE and JUMP_COVERAGE set to 45, CLOSE_UNIPATH_GAPS set to FALSE and HAPLOIDIFY set to TRUE. The output of SSPACE-LR was further scaffolded using SSPACE (version 3.0)[65,68] with the two sets of quality-trimmed mate-pair reads and the following options: -x 0 -m 32 -o 20 -k 5 -a 0.70 -n 15 -p 0 -v 0 -z 0 -g 0 -T 32 -S 0. Quality trimming of the 4-kbp mate-pair reads was done using HTQC[69]. Quality trimming of the 8-kbp mate-pair reads was done using cutadapt[70] and Trimmomatic[71]. Scaffolding with SSPACE-LR was repeated before gaps were filled with PacBio reads using PBJelly (version 15.8.24)[72] with -t 1000 -w 4000 options at the assembly step. All filtered reads without error correction were used for the gap filling with PBJelly. Additional scaffolding steps with SSPACE and

SSPACE-LR were carried out after the gap filling. Final scaffolding was performed using L_RNA_scaffolder[73] combined with the de novo transcriptome assembly (see below). Finally, gaps were filled using Sealer (version 1.9.0)[74] and quality-trimmed 250-bp paired-end reads with -P 100 and -B 5000 options by scanning k-mer sizes from 96 through 86. Quality trimming of the 250-bp paired-end reads was done using Trimmomatic[71]. Assembly statistics at each step of the assembly pipeline are shown in Supplementary Table 3. Scaffolds larger than 2 kbp were used to calculate the final assembly statistics in Supplementary Table 1.

Isolation of mRNA, library preparation and de novo transcriptome sequencing. DNA/RNA was extracted from samples using a phenol/chloroform protocol, and total RNA was isolated using a clean-up step with TRI reagent (Sigma-Aldrich). Details of the protocol are descried in the Supplementary Methods. The concentration and integrity of each RNA extraction was verified using a 2100 Bioanalyzer (Agilent). Total RNA was converted into tagged complementary DNA libraries using the TruSeq RNA Sample Preparation Kit v2 (Illumina) according to the manufacturer's protocol. Libraries were sequenced using an Illumina HiSeq 2000. We began by running 1 polyp sample on 1 lane with 100-nucleotide paired-end sequencing. After vetting the results, we performed additional 100-nucleotide paired-end sequencing on samples across the life cycle. These paired-end data sets were used for the de novo transcriptome assembly. Additional biological replicates were sequenced using 50-nucleotide single-end reads. Details about each sample and the relevant NCBI Sequence Read Archive accessions are provided in Supplementary Table 2.

Gene prediction and annotation. The annotation pipeline is described in detail in the Supplementary Methods and illustrated in Supplementary Fig. 1. Briefly, de novo transcriptome assembly was performed using Trinity[75], and this data was passed to PASA[76]. Ab initio predictions were performed using GeneMark-ES[77], glimmerHMM[78] and the AUGUSTUS web server[79] with default settings. Trinity models and the Uniprot Swissprot protein data set were mapped to the genome using exonerate[80] and GMAP[81]. All gene models were passed to EVidenceModeler[76] to create a weighted consensus gene structure data set, and the weighted models were passed back into PASA to create a final set of predictions[76].

Following gene modelling, the results went through an annotation pipeline that included the following analyses: (1) BLASTp of protein models against the Uniprot Swissprot data set, (2) BLASTx of transcript models against the Uniprot Swissprot data set and (3) protein domain identification using HMMER and the Pfam-A database[82,83]. Gene models were rejected if they lacked a protein model and Uniprot annotation and had less than ten total reads mapped from the RNA-seq analyses (described below). This resulted in a final count of 29,964 vetted gene models. An annotation report from this pipeline is included in the Supplementary Data. The gene annotations described above were used to create the tables comparing genes with conserved transcription-factor domains (Supplementary Table 8) and signalling molecules (Supplementary Table 9). Basic statistics on the gene models are provided in Supplementary Table 4.

Test for trans-spliced leader additions. Because the gene models are built off of the genomic backbone, we would not anticipate finding trans-spliced leader additions in this data. We instead used the de novo mRNA models, which were assembled by Trinity using 100-bp paired-end reads (see above). We performed two tests to look for conserved leader sequences. First, we used BLASTn to query all known *Clytia*[16] and *Hydra*[17] trans-spliced leader sequences against the Trinity mRNA models. After finding no hits, we truncated all Trinity mRNA models to the first 100 bp, and then performed an all-versus-all BLASTn analysis with an e-value cut-off of 10×10^{-5}. Only one pair of unrelated mRNA models (that is, not sharing the same cluster and/or gene identity in the Trinity output) shared a conserved region in this analysis. We therefore conclude that there is no evidence in our data for trans-spliced leader addition in *Aurelia*.

RNA-seq analysis. We used a genome-guided approach to RNA-seq. First, raw reads were aligned to the *Aurelia* genome using Hisat-2[84]. For paired-end data sets, only the first 50 nucleotides from the forward reads were used. Gene counts were then estimated with the StringTie package[85]. Following vetting of the data sets (Supplementary Fig. 7), differential gene expression was calculated using the EdgeR package[86]. Only vetted genes were included in the analysis. Differentially expressed genes were identified based on a false-discovery rate adjusted *P* value of 0.05, and a minimum fourfold change in expression in at least 1 life stage comparison. The StringTie count matrix used for EdgeR is provided in the Supplementary Data.

STRING analysis. For STRING analysis, all differentially expressed genes from *Aurelia* were queried against the predicted proteins for *Drosophila* (Uniprot identity: UP000000803) and *Homo* (Uniprot identity: UP000005640) using BLASTx (with a minimum e-value of 10×10^{-5}). The top BLAST hits were used to batch submit queries in the 'Multiple Proteins' section of the STRING v10 web server[39].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The genome assembly, as well as raw reads underlying the genomic and transcriptomic sequencing, are deposited in NCBI under BioProject PRJNA490213. A genome browser is also hosted at www.DavidAdlerGold.com/jellyfish. The Supplementary Data contain relevant input, intermediate and output data from all bioinformatics analyses performed in this paper. Annotations of the *Aurelia* gene models are provided in the Supplementary Data.

Received: 3 May 2017; Accepted: 12 October 2018;
Published online: 03 December 2018

References

1. Simion, P. et al. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr. Biol.* **27**, 958–967 (2017).
2. Whelan, N. V. et al. Ctenophore relationships and their placement as the sister group to all other animals. *Nat. Ecol. Evol.* **1**, 1737–1746 (2017).
3. Han, J. et al. The earliest pelagic jellyfish with rhopalia from Cambrian Chengjiang Lagerstätte. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **449**, 166–173 (2016).
4. Cartwright, P. et al. Exceptionally preserved jellyfishes from the Middle Cambrian. *PLoS ONE* **2**, e1121 (2007).
5. Putnam, N. H. et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
6. Baumgarten, S. et al. The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proc. Natl Acad. Sci. USA* **112**, 11893–11898 (2015).
7. Shinzato, C. et al. Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320–323 (2011).
8. Chapman, J. A. et al. The dynamic genome of *Hydra*. *Nature* **464**, 592–596 (2010).
9. Dawson, M. N. & Jacobs, D. K. Molecular evidence for cryptic species of *Aurelia aurita* (Cnidaria, Scyphozoa). *Biol. Bull.* **200**, 92–96 (2001).
10. Zapata, F. et al. Phylogenomic analyses support traditional relationships within Cnidaria. *PLoS ONE* **10**, e0139068 (2015).
11. Nakanishi, N., Yuan, D., Hartenstein, V. & Jacobs, D. K. Evolutionary origin of rhopalia: insights from cellular-level analyses of Otx and POU expression patterns in the developing rhopalial nervous system. *Evol. Dev.* **12**, 404–415 (2010).
12. Katsuki, T. & Greenspan, R. J. Jellyfish nervous systems. *Curr. Biol.* **23**, R592–R594 (2013).
13. Leclère, L. et al. The genome of the jellyfish *Clytia hemisphaerica* and the evolution of the cnidarian life-cycle. Preprint at https://www.biorxiv.org/content/early/2018/07/20/369959 (2018).
14. Goldberg, R. B. et al. DNA sequence organization in the genomes of five marine invertebrates. *Chromosoma* **51**, 225–251 (1975).
15. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
16. Derelle, R. et al. Convergent origins and rapid evolution of spliced leader trans-splicing in metazoa: insights from the ctenophora and hydrozoa. *RNA* **16**, 696–707 (2010).
17. Stover, N. A. & Steele, R. E. Trans-spliced leader addition to mRNAs in a cnidarian. *Proc. Natl Acad. Sci. USA* **98**, 5693–5698 (2001).
18. Bellis, E. S., Howe, D. K. & Denver, D. R. Genome-wide polymorphism and signatures of selection in the symbiotic sea anemone *Aiptasia*. *BMC Genomics* **17**, 160 (2016).
19. Fernandez-Valverde, S. L., Calcino, A. D. & Degnan, B. M. Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge *Amphimedon queenslandica*. *BMC Genomics* **16**, 387 (2015).
20. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
21. Levin, M. et al. The mid-developmental transition and the evolution of animal body plans. *Nature* **531**, 637–641 (2016).
22. Park, E. et al. Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. *Mol. Phylogenet. Evol.* **62**, 329–345 (2012).
23. Fuchs, B. et al. Regulation of polyp-to-jellyfish transition in *Aurelia aurita*. *Curr. Biol.* **24**, 263–273 (2014).
24. Brekhman, V., Malik, A., Haas, B., Sher, N. & Lotan, T. Transcriptome profiling of the dynamic life cycle of the scypohozoan jellyfish *Aurelia aurita*. *BMC Genomics* **16**, 74 (2015).
25. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
26. Boehm, A.-M. et al. FoxO is a critical regulator of stem cell maintenance in immortal Hydra. *Proc. Natl Acad. Sci. USA* **109**, 19697–19702 (2012).
27. Bridge, D. et al. FoxO and stress responses in the cnidarian *Hydra vulgaris*. *PLoS ONE* **5**, e11686 (2010).

28. Coutts, A. S., Weston, L. & La Thangue, N. B. A transcription co-factor integrates cell adhesion and motility with the p53 response. *Proc. Natl Acad. Sci. USA* **106**, 19872–19877 (2009).
29. Hemmrich, G., Miller, D. J. & Bosch, T. C. The evolution of immunity: a low-life perspective. *Trends Immunol.* **28**, 449–454 (2007).
30. Srivastava, M. et al. The *Trichoplax* genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).
31. Ryan, J. F., Pang, K., Mullikin, J. C., Martindale, M. Q. & Baxevanis, A. D. The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *EvoDevo* **1**, 9 (2010).
32. Ryan, J. F. et al. The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: evidence from the starlet sea anemone, *Nematostella vectensis*. *Genome Biol.* **7**, R64 (2006).
33. Srivastava, M. et al. The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010).
34. Chourrout, D. et al. Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* **442**, 684–687 (2006).
35. Consortium, U. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2016).
36. Nakanishi, N., Hartenstein, V. & Jacobs, D. K. Development of the rhopalial nervous system in *Aurelia* sp. 1 (Cnidaria, Scyphozoa). *Dev. Genes Evol.* **219**, 301–317 (2009).
37. Helm, R. R., Tiozzo, S., Lilley, M. K., Lombard, F. & Dunn, C. W. Comparative muscle development of scyphozoan jellyfish with simple and complex life cycles. *EvoDevo* **6**, 11 (2015).
38. Gold, D. A. et al. Structural and developmental disparity in the tentacles of the moon jellyfish *Aurelia* sp. 1. *PLoS ONE* **10**, e0134741 (2015).
39. Szklarczyk, D. et al. STRINGv10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2014).
40. Gehring, W. J. The evolution of vision. *Wiley Interdiscip. Rev. Dev. Biol.* **3**, 1–40 (2014).
41. Albert, D. J. What's on the mind of a jellyfish? A review of behavioural observations on *Aurelia* sp. jellyfish. *Neurosci. Biobehav. Rev.* **35**, 474–482 (2011).
42. Nakanishi, N., Camara, A. C., Yuan, D. C., Gold, D. A. & Jacobs, D. K. Gene expression data from the moon jelly, *Aurelia*, provide insights into the evolution of the combinatorial code controlling animal sense organ development. *PLoS ONE* **10**, e0132544 (2015).
43. Jacobs, D. K. et al. in *Key Transitiitions in Animal Evolution* (eds Schieirwater, B. & DeSalle, R.) Ch. 8, 175–193 (CRC Press, Boca Raton, 2010).
44. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* **9**, 868–882 (2008).
45. Kraus, J. E., Fredman, D., Wang, W., Khalturin, K. & Technau, U. Adoption of conserved developmental genes in development and origin of the medusa body plan. *EvoDevo* **6**, 23 (2015).
46. Sanders, S. M. & Cartwright, P. Patterns of Wnt signaling in the life cycle of *Podocoryna carnea* and its implications for medusae evolution in Hydrozoa (Cnidaria). *Evol. Dev.* **17**, 325–336 (2015).
47. Suga, H. et al. Flexibly deployed Pax genes in eye development at the early evolution of animals demonstrated by studies on a hydrozoan jellyfish. *Proc. Natl Acad. Sci. USA* **107**, 14263–14268 (2010).
48. Suga, H., Schmid, V. & Gehring, W. J. Evolution and functional diversity of jellyfish opsins. *Curr. Biol.* **18**, 51–55 (2008).
49. Graziussi, D. F., Suga, H., Schmid, V. & Gehring, W. J. The "*Eyes absent*" (*eya*) gene in the eye-bearing hydrozoan jellyfish *Cladonema radiatum*: conservation of the retinal determination network. *J. Exp. Zool. B Mol. Dev. Evol.* **318**, 257–267 (2012).
50. Artigas, G. Q. et al. A gonad-expressed opsin mediates light-induced spawning in the jellyfish *Clytia*. *eLife* **7**, e29555 (2018).
51. Liegertová, M. et al. Cubozoan genome illuminates functional diversification of opsins and photoreceptor evolution. *Sci. Rep.* **5**, 11885 (2015).
52. Sanders, S. M. & Cartwright, P. Interspecific differential expression analysis of RNA-seq data yields insight into life cycle variation in hydractiniid hydrozoans. *Genome Biol. Evol.* **7**, 2417–2431 (2015).
53. Ames, C. L., Ryan, J. F., Bely, A. E., Cartwright, P. & Collins, A. G. A new transcriptome and transcriptome profiling of adult and larval tissue in the box jellyfish *Alatina alata*: an emerging model for studying venom, vision and sex. *BMC Genomics* **17**, 650 (2016).
54. Collins, A. G., Cartwright, P., McFadden, C. S. & Schierwater, B. Phylogenetic context and basal metazoan model systems. *Integr. Comp. Biol.* **45**, 585–594 (2005).
55. Erwin, D. H. et al. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* **334**, 1091–1097 (2011).
56. Hyman, L. H. *The Invertebrates: Protozoa Through Ctenophora* (McGraw-Hill, New York, 1940).
57. Marques, A. C. & Collins, A. G. Cladistic analysis of medusozoa and cnidarian evolution. *Invertebr. Biol.* **123**, 23–42 (2004).
58. Bridge, D., Cunningham, C. W., Schierwater, B., Desalle, R. O. B. & Buss, L. W. Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. *Proc. Natl Acad. Sci. USA* **89**, 8750–8753 (1992).
59. Moroz, L. L. et al. The ctenophore genome and the evolutionary origins of neural systems. *Nature* **510**, 109–114 (2014).
60. Ryan, J. F. et al. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* **342**, 1242592 (2013).
61. Gold, D. A. Life in changing fluids: a critical appraisal of swimming animals before the Cambrian. *Integr. Comp. Biol.* **58**, 677–687 (2018).
62. Condon, R. H. et al. Jellyfish blooms result in a major microbial respiratory sink of carbon in marine systems. *Proc. Natl Acad. Sci. USA* **108**, 10225–10230 (2011).
63. Abrams, M. J., Basinger, T., Yuan, W., Guo, C.-L. & Goentoro, L. Self-repairing symmetry in jellyfish through mechanically driven reorganization. *Proc. Natl Acad. Sci. USA* **112**, E3365–E3373 (2015).
64. Boetzer, M. & Pirovano, W. SSPACE-Long Read: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**, 211 (2014).
65. Hackl, T., Hedrich, R., Schultz, J. & Förster, F. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**, 3004–3011 (2014).
66. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
67. Gnerre, S. et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl Acad. Sci. USA* **108**, 1513–1518 (2011).
68. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2010).
69. Yang, X. et al. HTQC: a fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics* **14**, 33 (2013).
70. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).
71. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
72. English, A. C., Salerno, W. J. & Reid, J. G. PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. *BMC Bioinformatics* **15**, 180 (2014).
73. Xue, W. et al. L_RNA_scaffolder: scaffolding genomes with transcripts. *BMC Genomics* **14**, 604 (2013).
74. Paulino, D. et al. Sealer: a scalable gap-closing application for finishing draft genomes. *BMC Bioinformatics* **16**, 230 (2015).
75. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
76. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
77. Lukashin, A. V. & Borodovsky, M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**, 1107–1115 (1998).
78. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
79. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309–W312 (2004).
80. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
81. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
82. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
83. Bateman, A. et al. The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
84. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
85. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
86. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
87. Technau, U. & Steele, R. E. Evolutionary crossroads in developmental biology: Cnidaria. *Development* **138**, 1447–1458 (2011).
88. Zhong, Y. & Holland, P. W. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol. Dev.* **13**, 567–568 (2011).
89. Larroux, C. et al. Genesis and expansion of metazoan transcription factor gene classes. *Mol. Biol. Evol.* **25**, 980–996 (2008).

Acknowledgements

We thank K. Kosik and N. Nakanishi for their insights during the development of this project; R. Warren for his advice on genome assembly strategy; V. Levesque and the Birch Aquarium at Scripps for providing *Aurelia* strains; and S. Johnson, D. Le, D. Lam,

and A. Hsu for technical assistance. D.A.G. gratefully acknowledges funding from a National Institutes of Health Training Grant in Genomic Analysis and Interpretation (T32HG002536) and a Cordes Postdoctoral Fellowship from the Division of Biology and Biological Engineering at Caltech. This work was also supported by grants from the W.M. Keck Foundation (R.J.G.), the Gordon and Betty Moore Foundation (R.J.G.), the DFG (T.H.), a fellowship from the Uehara Memorial Foundation (T.K.) and the NASA Astrobiology Institute–Foundations of Complex Life: Evolution, Preservation and Detection on Earth and Beyond (D.K.J.).

Author contributions

R.J.G. and T.K. sequenced and assembled the genome with input from R.E.S. M.R. contributed 100-bp paired-end reads, and D.I. and T.H. provided mate-pair reads with 4-kbp inserts. X.Y. and Y.L. worked on the initial error correction of PacBio reads. D.K.J. and D.A.G. oversaw transcriptome sequencing and assembly. D.A.G. performed downstream analyses of genome annotation with input from T.K., R.J.G., R.E.S. and D.K.J. D.A.G. designed the figures and drafted the manuscript. All authors reviewed and approved the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0719-8.

Correspondence and requests for materials should be addressed to D.A.G. or T.K. or R.J.G.

nature research

Corresponding author(s): David A. Gold
Takeo Katuski
Ralph J. Greenspan

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	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 all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted Give P values as exact values whenever suitable.
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☒	☐	Clearly defined error bars State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection, other than the software built into the PaBio and Illumina next-generation sequencers.
Data analysis	Open-source, publicly-available software used in this study includes: Discovar de novo; FLASh; ALLPATHS-G; Proovread; SSPACE; PBJelly; L_RNA_Scaffolder; Sealer; Trinity; PASA; GMAP; GeneMark; AUGUSTUS; GlimmerHMM; Exonerate; EvidenceModeler; HiSat2; StringTie; RepeatScout; RepeatMasker; OrthoFinder; PhyML; and BEAST. All programs used in this study are appropriately cited in the manuscript, with software version numbers when applicable.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The genome assembly, as well as raw reads underlying the genomic and transcriptomic sequencing, are deposited in NCBI under BioProject PRJNA490213. A genome browser is also hosted at www.DavidAdlerGold.com/jellyfish. Additional File 1 contains relevant input, intermediate, and output data from all bioinformatics analyses performed in this paper. Annotation of the Aurelia gene models is provided as Additional File 2.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This paper describes the sequencing of the moon jellyfish (Aurelia) genome. We use genomic and transcriptomic data to compare the Aurelia genome to genomes of morphologically simpler relatives: the sea anemones, corals, and hydras.
Research sample	Samples for DNA and RNA sequencing were taken from the "species 1" isolate of Aurelia. We chose this population because it is native to the California coastline. This is the state that the primary researchers on this project work in.
Sampling strategy	For genomic DNA sequencing, all DNA was extracted from a clonal population of animals. For RNA sequencing, animals were pooled together to get sufficient RNA yields. Six time points were analyzed for RNA-Seq analysis, and 2-3 biological replicates were included for each time point. The process is described in greater detail in the supplemental Materials and Methods.
Data collection	Data was collected by T.K. and D.A.G. using standard next-generation sequencers (PacBio, Illumina). The process is described in greater detail in the supplemental Materials and Methods.
Timing and spatial scale	All DNA and RNA samples were collected within a 1-year window. An initial RNA sample was collected to vet the extraction and sequencing protocol. Following vetting, all additional RNA samples were collected. There was approximately six months between the first RNA sample and the additional samples. All additional samples were collected over a ~2 month period.
Data exclusions	Several RNA samples were excluded from the final analysis because they failed to meet standard quality control expectations (number of mapped reads, poor quality scores at 3' ends). These were not included in the gene models or downstream analyses.
Reproducibility	Reproducibility between biological replicates was visualized using multidimensional scaling plots. These are provided in Figure S7 of the manuscript.
Randomization	Randomization was not relevant to this study, as we were not testing for differences between populations.
Blinding	Blinding was not relevant to this study, as we were not testing for differences between populations.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	tyrosinated tubulin (Sigma, cat#T9028). This is reported in the manuscript (Figure 1 legend)
Validation	This antibody has previously been validated in Aurelia (and many marine invertebrates) in previous studies. For example, see Yuan et al. (2008) Dev. genes and evol. 218(10), 525–539; Nakanishi e al. (2009) Dev. genes and evol., 219(6), 301–317.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Our animals are Aurelia aurita. "species 1" strain. They were originally caught in the wild and raised in laboratories for multiple years. The clonal population that the genome was sampled from is now raised at UC San Diego, UCLA, UC Davis, and Caltech. Voucher specimens have also been submitted to the Smithsonian Institution.
Wild animals	Our animals are Aurelia aurita. "species 1" strain, caught in the wild and raised in laboratories for multiple years.
Field-collected samples	Our animals are Aurelia aurita. "species 1" strain, caught in the wild and raised in laboratories for multiple years.