Krüppel-like factor/specificity protein evolution in the Spiralia and the implications for cephalopod visual system novelties

Kyle J. McCulloch1,2 and Kristen M. Koenig1,2

1Department of Organismic and Evolutionary Biology, Harvard University Cambridge, MA 02138, USA
2John Harvard Distinguished Science Fellows, Harvard University, Cambridge, MA 02138, USA

The cephalopod visual system is an exquisite example of convergence in biological complexity. However, we have little understanding of the genetic and molecular mechanisms underpinning its elaboration. The generation of new genetic material is considered a significant contributor to the evolution of biological novelty. We sought to understand if this mechanism may be contributing to cephalopod-specific visual system novelties. Specifically, we identified duplications in the Krüppel-like factor/specificity protein (KLF/SP) sub-family of C2H2 zinc-finger transcription factors in the squid Doryteuthis pealeii. We cloned and analysed gene expression of the KLF/SP family, including two paralogs of the DpSP6-9a gene. These duplicates showed overlapping expression domains but one paralog showed unique expression in the developing squid lens, suggesting a neofunctionalization of DpSP6-9a. To better understand this neofunctionalization, we performed a thorough phylogenetic analysis of SP6-9 orthologues in the Spiralia. We find multiple duplications and losses of the SP6-9 gene throughout spiralian lineages and at least one cephalopod-specific duplication. This work supports the hypothesis that gene duplication and neofunctionalization contribute to novel traits like the cephalopod image-forming eye and to the diversity found within Spiralia.

1. Introduction

The generation of new genetic material subject to mutation is considered a significant contributor to the evolution of biological novelty [1]. This can include large genomic expansions and rearrangements, gene duplications and de novo gene emergence, cis-regulatory element expansion or exon shuffling [2–4]. However, a lack of sequence data has made it difficult to assess the importance of this mechanism in most metazoan species until recently. Cephalopods, a group of molluscs that include squid, cuttlefish, octopus, and nautilus, have recently benefited from genome and transcriptome sequencing and are exceptional subjects for the study of morphological novelty because they have evolved a diversity of complex, class-specific traits [5–12].

Cephalopod novelties include many features traditionally associated with vertebrates (closed circulatory system, single-chambered eyes) as well as some defining to themselves (ink decoys and textured, camouflaging skin). Little is known about the molecular or developmental basis of these traits, including whether the genetic and genomic changes that underpin these novelties are similar to the mechanisms found in better-studied species. Cephalopod genomes are large, ranging between two and five gigabases [13]. Sequencing the octopus genome found evidence of large-scale transposon-mediated gene family expansions [10]. One of these expansions is the large C2H2 zinc-finger superfamily of proteins, with over double the number of homologues found
in vertebrates. This is consistent in the genome of the bobtail squid, *Euprymna scolopes*, and in the transcriptome of the longfin squid, *Doryteuthis pealeii* [6,8]. To understand the contribution of C2H2 zinc-finger protein family expansion to the evolution of cephalopod novelty, we have chosen to focus on a sub-family with well-understood function in vertebrates and *Drosophila*: the KLF (Krüppel-like factor) and SP (specificity protein) genes [14–17]. KLF/SP genes comprise a family of related transcription factors found within the Holozoa characterized by a conserved triple C2H2 zinc-finger DNA-binding domain at the C-terminus. The SP proteins are restricted to the Metazoa and are defined by a Buttonhead box (Btd), a putative transactivating domain, and usually an SP box located at the N-terminus (figure 1a; electronic supplementary material, figure S1) [18]. SP1 was the first eukaryotic transcription factor identified. Since the discovery of SP1, KLF/SP proteins have been implicated in diverse developmental, cellular, and homeostatic processes [19,20]. They have known roles as both transcriptional activators and repressors in neural development, angiogenesis, osteogenesis, muscle, digestive, and renal physiology, as well as in cancer progression [19–26]. The KLF/SP triple C2H2 domain binds to GC/GT boxes in DNA and its biochemical function is conserved across distantly related species [27]. Partial functional redundancy is often observed in overlapping paralog expression, suggesting that the context of expression is important for KLF/SP functional diversity rather than molecular differences [28]. There are 17 KLF and 9 SP genes in mice and 5 KLF and 3 SP genes in *Drosophila* with variable numbers identified in other vertebrates and ecdysozoans [29,30]. However, the phylogeny and function of this group of genes is poorly understood in the Spiralia. Analysis of previously published RNA-seq data in the squid *D. pealeii* [6] shows an enrichment of KLF/SP family members in the developing visual system, including the eye and optic lobe tissues, where image processing occurs (figure 1b) [31]. The visual system is a compelling subject for the study of biological novelty because it demonstrates significant innovations in coleoid, or soft-bodied, cephalopods [32–34].

Cephalopods are one of four groups of animals, including vertebrates, pancrustaceans, and arachnids, that have independently evolved a highly acute visual system [35]. Coleoid cephalopod eye morphology is physically similar to the vertebrate eye, composed of a single chamber with a lens and a cup-shaped retina (electronic supplementary material, figure S2). This complex organ evolved within Cephalopoda and is convergent with other camera-type eyes. The cephalopod lens is a lineage-specific morphological novelty that allows for high spatial resolution by precisely focusing light on the retina [36–38]. The cephalopod eyes develop from two placodes on either side of the embryo early in development. These placodes are then internalized by a lip of cells to form the optic vesicles. The posterior of the optic vesicle will generate the retina and the anterior will form the lens and anterior segment of the eye. At hatching, the neural retina is composed of two morphologically identifiable cell types: photoreceptors, which synapse directly on the optic lobe, and support cells. The anterior segment of the eye is composed of multiple populations of lens-generating or lentigenic cells. The lens is segmented into an anterior lens segment and a posterior lens segment. The lentigenic cells are arranged circumferentially around the lens and selectively contribute to each segment (electronic supplementary material, figure S2) [6,39,40]. Each population of lentigenic cells extends long processes that wrap around each other to form the lens fibres [39,40]. Gene expression studies and staged RNA-seq data of the developing eye and optic lobe have been published but we have little understanding of the molecular contributors to lens development in cephalopods [6,34,39,41–47].

Our goal in this study is to illuminate if gene duplications in the KLF/SP transcription factor family may be associated...
with visual system innovations in the cephalopod. Here, we identify duplications of the KLF/SP family of genes specific to cephalopods by generating a new phylogeny using recent genomic and transcriptomic data across Spiralia. To better understand the evolutionary history of one of these cephalopod duplications we focused on the SP6-9 homologues, which are enriched in the RNA-seq time course of the eye and optic lobe development. Our results show that the SP6-9 gene has undergone multiple duplications and losses within Spiralia including a cephalopod-specific duplication. We cloned and performed in situ hybridization on members of the KLF/SP family to assess their expression in the squid visual system. Our study showed lens-specific expression of a single paralog of the DpSP6-9 gene identifying a potential neo-functionalization. This work supports the hypothesis that gene duplication plays an important role in the evolution of novelty and diversity found in the Spiralia.

2. Material and methods

(a) Animal husbandry
Egg sacs were obtained from the Marine Biological Laboratory, Woods Hole, Massachusetts. Embryos were kept in flowing seawater at the MBL or at 20°C in artificial sea water.

(b) Cloning and in situ hybridization
Genes were identified using a previously assembled and published transcriptome [6]. RNA was extracted from pooled stages of embryos using Trizol. cDNA libraries were generated usingScript according to the manufacturer's instructions (Bio-Rad). Primers used for cloning gene fragments are found in electronic supplementary material, table S1. PCR product size was confirmed by electrophoresis and then cloned into pGEM-T Easy and Sanger sequenced (Promega). Sense and antisense probes were generated from plasmid using digoxigenin-labelled ribonucleoside tri-phosphate (rNTPs) (Roche). In situ hybridization was performed as previously described [6]. Embryos were embedded and cryosectioned as previously described [6]. Whole-mount embryos were imaged using a Zeiss Axioskop 2.

(c) Phylogenetic analysis
The metazoan KLF/SP family member trees were generated from amino acid sequences using previous alignments [29,30], and NCBI BLAST searches in additional spiralians and basal deuterosomes. We required that sequences had three C2H2 domains with the following 88 amino acid architecture: C-X4-C-X12-H-X7-H-X4-C-X4-C-X12-H-X7-H-X4-C-X4-C-X12-H-X7-H, where X can be any amino acid. Additionally, we confirmed the presence of a conserved aspartic acid residue at position 44 (D44) [30]. Non-metazoan, opisthokont KLF sequences were used as outgroups. We aligned sequences in Muscle and implemented MrBayes as described above, for 565,000 generations until the average standard deviation of split frequencies was consistent at 0.019.

We identified SP6-9 sequences for our spiralian tree by reciprocal BLAST of available spiralian transcriptomes and genomes. We used full-length amino acid sequences for the SP6-9 trees when possible and used SP1-5 sequences as outgroups. We allowed the RAxML software to identify the best-scoring protein substitution model (PROTGAMMAAUTO) for our dataset and the JTT model with fixed base frequencies under the gamma model of rate heterogeneity was used. RAxML and MrBayes were otherwise implemented as above on the spiralian SP6-9 alignment. In the Bayesian analysis, the average standard deviation of split frequencies reached 0.096 after 10 million generations.

We built the cephalopod-only SP6-9 tree using the cephalopod amino acid sequences from our spiralian tree with vertebrate and D. pealeii SP1-5 sequences as outgroups. We aligned sequences in Muscle and implemented MrBayes as described above, for 565,000 generations until the average standard deviation of split frequencies was consistent at 0.019.

For the Mollusca and Annelida SP6-9 only trees, we trimmed, filtered, and assembled basally branching annelid and mollusc transcriptomes according to the Harvard FAS Informatics best practices, a publicly available pipeline found here: https://informatics.fas.harvard.edu/best-practices-for-de-novo-transcriptome-assembly-with-trinity.html. We used reciprocal BLAST to identify potential SP6-9 orthologues, aligned only subgroup specific sequences, and implemented MrBayes as above, with 4 chains in each of 2 runs, and 4 million generations. Alignment and nexus tree files may be found in supplementary data (electronic supplementary material, data S9–S12).

(d) Analysis of SP6-9 synteny
Using publicly available spiralian genomes, we identified genomic coordinates and intron/exon boundaries of SP6-9 paralogs. We used all available cephalopod genomes. To assess synteny in other spiralian, we used chromosome-level assemblies with the exception of annelids which lack a chromosome-level assembly [51].

(e) Character mapping duplications and losses of SP6-9
Cartoon cladograms were made by manually combining the most current hypotheses of relationships among Spiralia, Annelida, and Mollusca [52–61]. For simplicity, we only include branches for which we have data. We then mapped gains and losses to find the most parsimonious hypothesis. In the cases of Rottia, Sedentaria, and Hysugastropoda, we override maximum parsimony in favour of the phylogenetic signal we see from our gene trees.

3. Results and discussion

(a) Analysis of Doryteuthis pealeii Krüppel-like factor/specificity protein homologs
We identified DpKLF/SP genes in a previously published D. pealeii transcriptome from reciprocal BLAST [6]. Time course developmental expression in the eye and optic lobe of these putative DpKLF/SP genes is shown in figure 1b [6]. Each DpKLF/SP amino acid sequence has the conserved triple C2H2 domain and the DpSP sequences have the requisite Btd box (electronic supplementary material, figure S1a). We also find previously identified transactivation and
repres sor domains in these D. pealeii sequences, including the SID, PV DLS, R2, and R3 rep res sor domains and the 9a TAD transactivating do main (ele c tronic supplementary ma terial, f igure S1b) [30].

To validate the putative DpKL F/SP genes from our dataset, we con structed phylogenetic trees with representative basal metazoan and bilaterian KLF/SP family mem bers (ele c tronic supplementary ma terial, table S2). Pre vious phylogenetic analyses either omitted or had lim ited sampling within the Spiralia [29,30]. This gene family presents a chal lenge for phylo genetic analysis due to the highly divergent N-terminal regions and the highly conserved but short C2H2 do mains [17,30,62]. To produce the alignment, we concatenated any conserved activator and repres sor motifs, the C2H2 do main, and Btd box. A summary of the Bayesian tree gen erated for this gene family is labelled and coloured by major subgroups supported by posterior probabilities (f igure 1c). ML meth ods never generated a tree with support for all previously identified KLF/SP groups (ele c tronic supplementary ma terial, data S2). In our Baysian tree, we find support for sub-families: KLF15, KLF1-8+12, KLF9/13, KLF10/11, SP1-4, and SP5-9. Within Spiralia, we find most taxa have one KLF15 rep resentative, four KLF1-8+12 members, a single KLF9/13 member, one SP1-4, and three SP5-9 mem bers.

Our tree supports subgroup clades but the addition of understudied taxa does not resolve 1:1 orthology for many sequences. We found one D. pealeii KLF15 and three KLF1-8+12 family members. Within the KLF1-8+12 sequences, previously identified jawed vertebrate subgroups KLF5, KLF6/7, and KLF3/8/12 are well-supported, but the vertebrate KLF1/2/4 group was not resolved [15,63,64]. We cannot confi dently assign orthology to many protostome KLF1-8+12 sequences. KLF6/7 is an excep tion and contains representatives from all metazoan lineages except Spiralia. KLF5 is not sup ported as a monophyletic group. We find a vertebrate KLF5 clade and an in vertebrate KLF5 clade, which includes D. pealeii. Addition ally, despite only finding support for a vertebrate KLF3/8/12 clade, which is defined by a conserved PV DLS do main, there are Drosophila and spiralian sequences with a PV DLS do main. Cephalopods have two KLF1-8+12 sequences and one KLF5. Aside from DpKLF5, the D. pealeii KLFs in this subgroup are too divergent to confidently assess orthology, so we name them DpKLF1-1 and DpKLF1-2.

KLF9/13 sequences have been previously identified in many metazoan taxa but only the vertebrate clade is supported in our tree (f igure 1c) [30,65]. Although not monophyletic, we hypothesize that the polytomies neighbouring the vertebrate clade are KLF9/13 which includes sequences that have been identified in previous phylogenies [30]. Interest ingly, spiralian and Nautilus are represented, but no other cephalopod has a KLF9/13 sequence, suggesting a loss in Coleo id cephalopods. KLF10/11 was previously reported to be a vertebrate-specific gene, but with the addition of more pro tostome taxa, we identify a highly sup ported monophyletic KLF10/11 group, with a cephalopod-specific duplication (f igure 1c). D. melanogaster does not possess a KLF10/11 gene, but orthologues are found in Tribolium, Linitus, Daphnia, and multiple spiralian sequences.

The SP family is strongly supported in the tree and the phy logeny shows two SP groups, SP1-4 and SP5-9, neither with strong support (f igure 1c). Protostomes, including cephalopods, have one SP1-4 family member, with the excep tion of D. pealeii, which is unique in having two. SP5 and SP6-9 groups were not resolved, unlike in other trees with less spiralian sampling. Spiralia have variable numbers of SPS-9 genes, typically between 3 and 4.

The addition of spiralian taxa to the KLF/SP gene family tree reveals a pattern of spiralian-specific gains and losses in nearly every major KLF/SP group. This complex evolutionary trajectory differs from the majority of ecdysozoans and basal deuterostomes which have not significantly expanded their KLF/SP repertoires (f igure 1c), and may be responsible for difficulties in assigning orthology [30,66]. As is the case for other gene families with a very small highly conserved region, to better resolve this difficult tree, it is necessary to focus our inquiry on specific gene families and to narrow our phylogenetic sampling where full-length sequences may help. We decided to focus on the SP6-9 group as it has undergone a duplication event, DpSP6-9a and DpSP6-9b, and is enriched in the eye and optic lobe RNA-seq time course.

(b) Phylogenetic analysis of the SP6-9 amino acid sequences in Spiralia

Our large KLF/SP tree did not have sufficient support to iden tify SP5 and SP6-9 as distinct groups. To confirm orthology of our cephalopod sequences, we generated spiralian-specific SP6-9 trees. This sub-family shares much more sequence iden tity within Spiralia, therefore, when possible, alignments of full-length amino acid sequences were made to construct ML and Bayesian trees (ele c tronic supplementary ma terial, f igure S3). Our Bayesian tree resolved SP5 as a monophyletic group within SP6-9 but our ML tree shows support for an SP6-9 group to the exclusion of SP1-5 (ele c tronic supplementary ma terial, f igure S3). Previous work supports the conclusion of our ML tree but we proceed with a cephalopod-specific tree to confirm the SP5 and SP6-9 clades (see section c). We recovered three cephalopod SP6-9 paralogs, although we did not find SP6-9c in the D. pealeii transcriptome. We discovered single orthologues in Brachiopoda, Gastrotricha, Nemertea, and aculiferan molluscs, two paralogs in Rotifera, Platyhel minthes, Bryozoa, Phoronida, Gastropoda, Scaphopoda, and Bivalvia, and one to three copies in Annelida. We find modest support for monophyletic clades of SP6-9 orthologues in each spiralian lineage, but no support for deeper phylogenetic relationships between SP6-9 orthologues.

(c) Expression analysis of Krüppel-like factor/specificity protein homologues in Doryteuthis pealeii

Little is known about the function of the KLF/SP genes in the Spiralia, so we cloned and performed in situ hybridization studies in D. pealeii on multiple members of the family, including KLF5, SP1-4, and SP5. We analysed gene expression at stages 19, 21, 23, 25, and 27 in whole-mount (ele c tronic supplementary ma terial, f igure S4-S6) [67].

KLF5 is expressed in yolk nuclei, in a salt and pepper pattern across the cerebral ganglion, in arm IV (tentacle), and the dorsal gill at stage 19 through 23. Later in develop ment (stage 25), KLF5 is still found in the arms and expands to the funnel and mantle tissue. At stage 27, the expanded epidermal expression is detected (ele c tronic supplementary ma terial, f igure S4). SP1-4a is found ubiqui tously expressed from stage 19 until stage 27, consistent with vertebrate and arthropod expression (ele c tronic
supplementary material, figure S5) [29] reviewed in [68]. The second member of the SP gene family assessed, SP5, has dynamic expression in D. pealeii including high expression in the optic lobe at stage 23 (electronic supplementary material, figure S6). In all stages, low expression is seen in the retina, and anterior segment expression is seen at stage 27.

(d) Paralog-specific expression of the DpSP6-9 genes

To understand the consequence of gene duplication, we also performed expression studies for both DpSP6-9 paralogs. DpSP6-9a and DpSP6-9b have significant overlap in their expression with notable differences. Both DpSP6-9a and DpSP6-9b show expression on the medial side of the placode at stages 19 and 21 (figures 2b,c,g,h and 3a–i). The expression is also seen in the gills and the limbs in D. pealeii (figures 2 and 3). SP6-9 is part of a limb gene regulatory network found in vertebrates and ecdysozoans [69–75]. Recent findings have shown SP6-9a (SP8/9a) expressed early in cuttlefish limb outgrowth [12]. Both paralogs are also expressed in the anterior chamber organ from stage 19 through to stage 27, a tissue currently hypothesized to control ocular pressure but without a well-understood function [76].

Despite their similar expression patterns, these paralogs do show differences in their domains. DpSP6-9a shows enriched expression in regions hypothesized to form the cerebral ganglion and the palliovisceral ganglion (figure 2c) and DpSP6-9b shows unique expression in the mantle at stage 23 until hatching (figure 2i–k). Major differences in expression are found in the developing eye. At stage 19, the expression of DpSP6-9a extends in all lip cells around the developing placode, while DpSP6-9b expression is restricted from the lip, with the exception of the most medial cells (figure 3a–e). At stage 21, the lip cells have fused in the anterior of the optic vesicle. At this stage, DpSP6-9a is expressed in the anterior segment and DpSP6-9b is significantly reduced in the anterior segment (figure 3a,f–i). This difference in expression is maintained through to stage 27. At stages close to hatching, DpSP6-9a is in the tertiary lentigenic cells in the anterior segment (figure 3f) [39,40].

This significant difference in expression suggests that SP6-9a plays a role in lens development while SP6-9b does not. Vertebrate and Drosophila homologues of SP6-9 do not have an association with visual system development but do show expression in the central nervous system [75,77,78]. However, SP6-9 expression in the retina may be ancestral in spiralian visual systems as SP6-9 is required for eye cup regeneration in the flatworm Schmidtea mediterranea in conjunction with the transcription factor DIs [79]. The evolution of novel expression domains is characteristic of neofunctionalization after duplication [80]. In finding this paralog-specific expression correlated with a cephalopod novelty (the lens), we wanted to further investigate this duplication event.

(e) Cephalopod SP6-9 sequence evolution and synteny

To understand the molecular changes in the cephalopod SP6-9 genes correlated with divergent expression and to confirm our SP6-9/SP5 orthology, we constructed a Bayesian tree with only cephalopod members (figure 4a). This tree shows support for separate SP5 and SP6-9 groups. It also shows SP6-9a and SP6-9b as sister groups to the exclusion of SP6-9c in all cephalopods. This relationship between paralogs includes Nautilus, which do not have a lens. In addition, a long branch leads to cephalopod SP6-9a, but within this group sequences are highly similar. This contrasts with relatively shorter branches leading to both SP6-9b and SP6-9c clades, but increased sequence divergence within each group across cephalopod lineages (figure 4a). The similarity of cephalopod SP6-9a sequences includes the lensless Nautilus, which does not express any SP6-9 homologue in its eye [43]. Together this suggests that coding sequence evolution has not altered the
ancestral molecular function of SP6-9a. Instead, SP6-9a may perform a similar biochemical function among cephalopods, but its novel recruitment to the anterior segment is likely a result of a cis-regulatory change. Wnt is a well-studied regulator of SP6-9 homologues in both vertebrates and arthropods, and may play a role regulating this new expression domain in cephalopods [72,73,81–83].

Defining relationships among SP6-9 orthologues and paralogs is difficult with phylogenetic methods alone, so we sought additional evidence to assess their evolutionary history. We evaluated gene architecture and synteny in spiralian SP6-9 paralogs (figure 4b, electronic supplementary material, table S3). We found multiple instances of spiralian SP6-9 genes that were intronless, including all cephalopod SP6-9b and SP6-9c sequences. In addition, cephalopods share synteny across all SP6-9 paralogs. This includes SP6-9b, SP6-9c, and SP6-9a in sequential order, in the same 5’ to 3’ direction, within 1.5 Mb of each other in all cephalopod genomes investigated so far. We also find that SP6-9a is the only paralog in cephalopods found to have introns. SP6-9 duplicates in other spiralian species are on the same chromosome, but do not show the obvious synteny found within cephalopods (figure 4b). These tandem, intronless paralogs suggest two duplication events via retrotransposition early in the cephalopod lineage and maintained in the genome [84]. Retrotransposition is an important contributor to insertion events and structural variation in both vertebrate and arthropod species [85,86]. This may be an unusual case in cephalopods; however, despite the fact that the octopus genome is enriched in transposon activity, recent analyses have shown retrotransposon-derived genomic elements only make up 2–8% of molluscan genomes, as opposed to 35–52% in vertebrates [10,51,87–89].

**Figure 3.** *DpSP6-9a* but not *DpSP6-9b* is highly expressed in lentigenic cells of the developing eye. (a) Schematics showing lateral views of the developing eye, anterior is left. At stage 19, the retina placode is exposed and being internalized by a surrounding lip of cells (arrows). The vesicle is closed at stage 21 and the lip cells generate the lens and anterior segment. ol, optic lobe; lip, lip of cells that internalizes the retinal placode; l, lens, r, retina; ac, anterior chamber organ; a, arms. (b–u) Lateral view and cryosections of *DpSP6-9a* and *DpSP6-9b*. White arrowheads highlight the region of lip and developing lentigenic cells. In sections, anterior of the animal is up, medial is left. Expression in the retinal placode can be seen for both genes, r, retina. At stage 19, *DpSP6-9a* is highly expressed on both sides of the lip as it closes (arrows), where *DpSP6-9b* is restricted medially (b–e). At stage 21, the anterior of the eye vesicle expresses *DpSP6-9a* with minimal expression of *DpSP6-9b* r, retina; as, anterior segment (f–i). Later *DpSP6-9a* is strongly expressed in the lentigenic cells. Again *DpSP6-9b* is reduced (j–u). Asterisk, panel (t) is a stage 28 embryo; scale bars, 200 µm.
SUPPLEMENTARY MATERIAL Figure S3. Cephalopod architecture is distinct from other spiralians. Cephalopod represent a recent genomic duplication or genome assembly error. Genomic backbones in black are all to the same scale, and zoomed-in gene lengths are all to represent posterior probabilities above 0.5 (white), 0.7 (grey), and 0.95 (black). (f) Character mapping duplications and losses of SP6-9
From our SP6-9 trees, assigning specific lineage duplications and losses is difficult without support from a species phylogeny. We wanted to clarify if these SP6-9 duplications were cephalopod-specific and therefore needed to better understand the relationship of SP6-9 paralogs across Spiralia. We first estimated the number of SP6-9 duplicates at the base of Mollusca and Annelida. Our spiralian SP6-9 trees did not give strong support for deeper nodes, so we assembled additional transcriptomes with a particular focus on basal branching annelid and molluscan lineages, and generated Bayesian SP6-9 trees for 1 : 1 orthologues of SP6-9 sequences. Instead two paralogs with the exception of Protobranchia which has one SP6-9 orthologue. This supports a lineage-specific loss, with the understanding that transcriptomic sampling may be incomplete (figure 5c, electronic supplementary material, figure S3, pink). Within Gastropoda, the Hysopogastropoda have two SP6-9 paralogs but they fall within a single well-supported gastropod SP6-9 clade (electronic supplementary material, figure S3). Genomic evidence confirms the loss of one gastropod paralog and duplication of gastropod SP6-9-1 within the Hysopogastropoda lineage (figure 5c, electronic supplementary material, figure S3, green). The cephalopod-specific SP6-9 duplication and its relationship with a novel lens sits within a broader dynamic pattern of gene gain and loss in Spiralia. Ultimately, further research in this major branch of the animal tree will shed light on whether these patterns are in part responsible for the vast morphological diversity found in this group.

4. Conclusion
The cephalopod eye and optic lobe together is a remarkable example of a sophisticated biological system and the lens is one of its most apparent novelties. In this study, we identified at least one cephalopod-specific gene duplication of the SP6-9
Figure 5. SP6-9 has a complex pattern of duplication across Spiralia and Cephalopoda. (a) The most parsimonious pattern of duplication events mapped onto a current spiralian phylogeny. At minimum, separate duplications occurred in Rotifera, Platyhelminthes, Annelida, the branch leading to Phoronida & Bryozoa, and Conchifera. Annelida and Mollusca subtrees are collapsed and detailed in (b) and (c). (b) Annelida: amino acid tree evidence indicates two duplications within Sedentaria after the split of Orbiniidae and Sabellidae, followed by a loss and subsequent duplication event in Clitellata. Basally branching annelids have a single paralog only, suggesting a single paralog at the base of Annelida. (c) Mollusca: this hypothesis suggests a duplication event following the split of Aculifera. At minimum, there was a duplication in Cephalopoda, a loss in Protobranchia and a loss and duplication in Hypogastropoda. Phylogenies based on [52–61]. In all panels: star, gene duplication; X, gene loss; images from Phylopic.org, or drawn by K. McCulloch, with credit for elaboration: the evolution of new lens-specific neofunctionalization. It opens the door to many mechanisms ing a cephalopod-specific gene duplication to a visual system cephalopod anterior segment. This is the first study connect-its robust expression in the lentigentic cells in the developing architecture, and phylogenetics. SP6-9a differs from SP6-9b in gene as supported by similar gene expression, synteny, gene architecture, and phylogenetics. SP6-9a differs from SP6-9b in its robust expression in the lentigentic cells in the developing cephalopod anterior segment. This is the first study connect-ing a cephalopod-specific gene duplication to a visual system neofunctionalization. It opens the door to many mechanisms for elaboration: the evolution of new lens-specific cis-regulatory targets, or even the redeployment of canonical regulatory networks in the anterior segment of the cephalopod eye. This study highlights the exciting opportunities within Spiralia to address fundamental questions underlying the evolution of novel phenotypes, including the consequences of gene duplication and loss, as well as changes in cis-regulation.

Ethics. The animals were treated in accordance with the European directive 86/609/EEC (D.Lgs.n. 26/2014) with the goal of replacement, reduction, refinement, and the minimization of animal suffering.

Data accessibility. Cloning primers, alignment files, tree files with branch lengths, species list, sequence accession numbers, and references supporting our phylogenetic dataset have been uploaded as part of the electronic supplementary material. Two full-length and nine partial Doryteuthis pealeii KLF/SIP coding sequences are deposited in Genbank with accession numbers MT118824–MT118834.

Authors’ contributions. K.M.K. and K.J.M. designed all the experiments. K.J.M. and K.M.K. performed experiments. K.J.M. performed phylogenetic analyses. K.M.K. and K.J.M. wrote the manuscript.

Competing interests. We declare we have no competing interests.

Funding. K.M.K. and K.J.M.’s work is supported by the Office of the NIH Director 1DP5OD023111-01, Whitman Center Research Award, Baxter Postdoctoral Fellowship Fund, Eugene and Milicent Bell Fellowship Fund in Tissue Engineering at the MBL, the Elisabet Samuelsson Director’s Discretionary Fund, and Harvard University.

Acknowledgements. We would like to thank the Koenig and Srivastava lab members for helpful discussions. We also acknowledge the support of the John Harvard Distinguished Science Fellows community. We also thank the MBL and the Marine Resources Center.

References

1. Ohno S. 1970 Evolution by gene duplication. Berlin, Germany: Springer.
2. Taylor JS, Raes J. 2004 Duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet. 38, 615–643. (doi:10.1146/annurev.genet.38.072902.092831)
3. Lynch M, Conery JS. 2000 The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155. (doi:10.1126/science.290.5494.1151)
4. Ritschard EA, Whitelaw B, Albertin CB, Cooke IR, Strugnell JM, Simakov O. 2019 Coupled genomic
13. Albertin CB et al. 2020 A draft genome sequence of the elusive giant squid, Architeuthis dux. GigaScience 9, giz152. (doi:10.1093/gigascience/giz152)

11. Zarrella I et al. 2016 Eye development and photoreceptor differentiation in the cephalopod Doryteuthis pealeii. Development 143, 3168–3181. (doi:10.1242/dev.134254)

20. Moore DL, Apara A, Goldberg JL. 2011 Krüppel-like transcription factors in the nervous system: novel players in neurite outgrowth and axon regeneration. Mol. Cell. Neurosci. 47, 233–243. (doi:10.1016/j.mcn.2011.05.005)

21. Safe S, Abdelrahim M. 2005 Stp transcription factor family and its role in cancer. Eur. J. Cancer 41, 2438–2448. (doi:10.1016/j.ejca.2005.08.006)

22. Kim C-K, He P, Bliokhkeyva AB, Yang VW. 2017 SP and KLF transcription factors in digestive physiology and diseases. Gastroenterology 152, 1845–1875. (doi:10.1053/j.gastro.2017.03.035)

23. Jiang W, Cai J, Xie D, Wang L. 2012 Sp/KLF family and tumor angio genesis in pancreatic cancer. Curr. Pharm. Des. 18, 2420–2431. (doi:10.2174/13811821292240)

24. Black AR, Black JD, Azizkhan-Clifford J. 2001 Sp1 and kruppel-like factor family of transcription factors in cell growth regulation and cancer. J. Cell. Physiol. 188, 143–160. (doi:10.1002/jcp.1111)

25. Rane MJ, Zhao Y, Cai L. 2019 Krüppel-like factors (KLFs) in renal physiology and disease. EbihMedicine 40, 743–750. (doi:10.1016/j.ebimi.2019.01.021)

26. Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, de Crombrugghe B. 2002 The novel zinc-finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. Cell 108, 17–29. (doi:10.1016/s0092-8674(01)00622-5)

27. Xie Y, Li X, Zhang X, Mei S, Li H, Urso A, Zhu S. 2019 SP/KLF transcription factors: bring in the family. Rev. Physiol. Biochem. Pharmacol. 184, 1121. (doi:10.1007/s00704-019-00388-x)

28. Ogura A, Yoshida M-A, Shigeno S, Ikeo K, Shigeno S, Moritaki T, Young JZ. 1991 Light has many meanings for cephalopods. Vis. Neurosci. 47, 1845–1875. (doi:10.1016/j.gastro.2009.03.035)

29. Schaeper ND, Prpic N-M, Wimmer EA. 2010 A clustered set of three Sp-family genes is ancestral in cephalopods. Sci. Rep. 10, 353–360. (doi:10.1038/ncomms1698)

30. Presnell JS, Schnitzler CE, Browne WE. 2015 KLF/SP transcription factor family evolution: diversification, and innovation in eukaryotes. Genome Biol. Evol. 7, 2289–2309. (doi:10.1093/gbe/evu141)

31. Young JZ.1974 The central nervous system of Loligo I. The optic lobe. Phil. Trans. R. Soc. Lond. B 267, 263–302. (doi:10.1098/rstb.1974.0002)

32. Packard A. 1972 Cephalopods and fish: the limits of convergence. Biol. Rev. 47, 241–307. (doi:10.1111/j.1469-185x.1972.tb00975.x)

33. Young JZ. 1991 Light has many meanings for cephalopods. Vis. Neurosci. 7, 1–12. (doi:10.1017/s0952523800010907)

34. Yoshida MA, Ogura A, Ikeo K, Shigeno S, Moritaki T, Winters GC, Kohn A, Moroz LL. 2015 Molecular evidence for convergence and parallelism in evolution of complex brains of cephalopod molluscs: insights from visual systems. Integr. Comp. Biol. 55, 1070–1083. (doi:10.1093/icb/icv049)

35. Land MF, Nilsson D-E. 2012 Animal eyes. Oxford, UK: Oxford University Press.

36. Sweeney AM, Haddock SHD, Johnsen S. 2007 Comparative visual acuity of coleoid cephalopods. Int. Comp. Biol. 47, 808–814. (doi:10.1038/icb.icm092)

37. Sweeney AM, Des Marais D, Ban Y-EA, Johnsen S. 2007 Evolution of graded refractive index in squid lenses. J. R. Soc. Interface, 4, 685–698. (doi:10.1098/rsif.2006.0210)

38. Jagger WS, Sands PJ. 1999 A wide-angle gradient index optical model of the crystalline lens and eye of the octopus. Vis. Res. 39, 2841–2852. (doi:10.1016/j.jcp.2009.01.021)

39. Arnold JM. 1967 Fine structure of the development of the cephalopod lens. J. Ultrastruct. Res. 17, 527–543. (doi:10.1016/0022-5223(70)90139-4)

40. West JA, Sivak JG, Doughty MJ. 1995 MSocroscopic evaluation of the crystalline lens of the squid (Loligo opalescens) during embryonic development. Exp. Eye. Res. 60, 19–35. (doi:10.1016/s0040-674x(00)00800-7)

41. Yoshida M-A, Yura K, Ogura A. 2014 Cephalopod eye evolution was modulated by the acquisition of Pax-6 splicing variants. Sci. Rep. 4, 4256. (doi:10.1038/srep04256)

42. Yoshida M-A, Ogura A. 2011 Genetic mechanisms involved in the evolution of the cephalopod camera eye revealed by transcriptomic and developmental studies. BMC Evol. Biol. 11, 180. (doi:10.1186/1471-2148-11-180)

43. Soussouls K, Ogura A, Tsions PA. 2013 Transcriptome analysis of Nautilus and pygmy squid developing eye provides insights in lens and eye evolution. PLoS ONE 8, e70854. (doi:10.1371/journal.pone.0070854)

44. Ogura A, Yoshida M-A, Moritaki T, Okuda Y, Sese J, Shimizu KX, Soussouls K, Tsions PA. 2013 Loss of the six3/6 controlling pathways might have resulted in pinhole-eye evolution in Nautilus. Sci. Rep. 3, 1432. (doi:10.1038/srep01432)

45. Buresi A, Baratte S, Da Silva C, Bonnana L. 2012 orthodenticle/otx ortholog expression in the anterior brain and eyes of Sepia officinalis (Mollusca, Cephalopoda). Gene Expr. Patterns 12, 109–116. (doi:10.1016/j.gep.2012.02.001)

46. Imaizome B, Andouche A, Basaglia Y, Lopez P, Bonnana-Pontielli L. 2017 Eye development in Sepia officinalis embryos: What the uncommon gene expression profiles tells us about eye evolution. Front. Physiol. 8, 613. (doi:10.3389/fphys.2017.00613)

47. Yoshida M-A, Shigeno S, Tsukeni K, Funya H. 2010 Squid vascular endothelial growth factor receptor: a shared molecular signature in the convergent evolution of closed circulatory systems. Evol. Dev. 12, 25–33. (doi:10.1111/j.1525-142X.2009.00388.x)

48. Edgar RC. 2004 MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 5, 113. (doi:10.1186/1471-2105-5-113)

49. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large evolutionary histories as signatures of organisinal innovations in cephalopods. Bioessys 41, 1900073. (doi:10.1002/bies.201900073)
important for erythroid maturation. *Dev. Biol.* **403**, 115–127. (doi:10.1016/j.ydbio.2015.05.015)

65. Pei J, Grishin NV. 2015 C2H2 zinc finger proteins of the SP/KLF, Wilms tumor, EGR, Huckebein, and Klumppfus families in metazoans and beyond. *Gene* **573**, 91–99. (doi:10.1016/j.gene.2015.07.031)

66. Parker-Katiaeae L et al. 2007 Identification of the imprinted KLF14 transcription factor underlying human-specific accelerated evolution. *PloS Genet.* **3**, e0030065. (doi:10.1371/journal.pgen.0030065)

67. Arnold JM. 1965 Normal embryonic stages of the squid, Loligo pealii (lesueur). *Biol. Bull.* **128**, 24–32. (doi:10.2307/1539386)

68. O’Connor L, Gilmour J, Bonifer C. 2016 The role of the ubiquitously expressed transcription factor Sp1 in tissue-specific transcriptional regulation and in disease. *Yale J. Biol. Med.* **89**, 513–525.

69. Córdoba S, Requena D, Jory A, Saiz A, Estella C. 2016 The evolutionarily conserved transcription factor Sp1 controls appendage growth through Notch signaling. *Development* **143**, 3623–3637. (doi:10.1242/dev.138375)

70. Bell SM, Schreiner CM, Waclaw RR, Campbell K, Bricaud O, Collazo A, Meng A, Lin S. 2004 Pur alpha transcription factors, regulate common and diverged functions of the zinc finger transcription factor Sp8/9 in allometric appendage growth in the cephalopod Octopus Vulgaris. *Dev. Biol.* **275**, 427–435. (doi:10.1016/j.ydbio.2015.03.041)

71. Schick F, Purnell BA, Wimmer EA, Jackle H. 1999 Common and diverged functions of the Drosophila gene pair D-Sp1 and buttonhead. *Mech. Dev.* **89**, 125–132. (doi:10.1016/S0925-4773(99)00215-4)

72. Boycott BB, Young JZ. 1956 Reactions to shape in Octopus Vulgaris Larvacean. *Proc. Zool. Soc. London* **126**, 491–547. (doi:10.1111/j.1096-3645.1956.tb04513.x)

73. Penberthy WT, Zhao C, Zhang Y, Jesse JR, Yang Z, Bricaid O, Collazo A, Meng A, Lin S. 2004 Pur alpha and Sp8 as opposing regulators of neural gata2 expression. *Dev. Biol.* **275**, 225–234. (doi:10.1016/j.ydbio.2004.08.007)

74. Gaborieau E, Hurtado-Chong A, Fernández M, Azim K, Rainteau O. 2018 A dual role for the transcription factor Sp8 in postnatal neurogenesis. *Sci. Rep.* **8**, 14560. (doi:10.1038/s41598-018-32314-6)

75. Lapan SW, Redden PW. 2011 dkr and sp9 control optic cup regeneration in a prototypic eye. *PloS Genet.* **7**, e1002226. (doi:10.1371/journal.pgen.1002226)

76. Li W-H, Yang J, Gu X. 2005 Expression divergence between duplicate genes. *Trends Genet.* **21**, 602–607. (doi:10.1016/j.tig.2005.08.006)

77. Talamillo A, Delgado J, Nakamura T, de-Vega S, Yoshitomi Y, Unda F, Birnmeier W, Yamada Y, Ros MA. 2010 Role of expression, a zinc-finger transcription factor, in limb development. *Dev. Biol.* **337**, 363–374. (doi:10.1016/j.ydbio.2009.11.007)

78. Setton EFW, Sharma PP. 2018 Cooption of an appendage-patterning gene cassette in the head segmentation of anamniotes. *Proc. Natl Acad. Sci. USA.* **115**, E3491–E3500. (doi:10.1073/pnas.1720131115)

79. Estella C, Rieckhof G, Calieja M, Morata G. 2003 The role of buttonhead and Sp1 in the development of the ventral imaginal discs of Drosophila. *Development* **130**, 5929–5941. (doi:10.1242/dev.00832)

80. Kaessmann H, Vinkenbosch N, Long M. 2009 RNA-based gene duplication: mechanistic and evolutionary insights. *Nat. Rev. Genet.* **10**, 19–31. (doi:10.1038/nrg2487)

81. Cordaux R, Batzer MA. 2009 The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **10**, 691–703. (doi:10.1038/nrg2640)

82. McCullers TJ, Steiniger M. 2017 Transposable elements in Drosophila. *Mob. Genet. Elements.* **7**, 1–18. (doi:10.1080/21592568.2017.1318201)

83. Zhang G et al. 2012 The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54. (doi:10.1038/nature11413)

84. Takeuchi T et al. 2012 Draft genome of the pearl oyster Pinctada fucata: a platform for understanding bivalve biology. *DNA Res.* **19**, 117–130. (doi:10.1093/dnare/dss005)

85. Yoshida M-A, Ishikura Y, Moritaki T, Shoguchi E, Shimizu KK, Sese J, Ogura A. 2011 Genome structure analysis of mussels revealed whole genome duplication and lineage specific repeat variation. *Gene* **483**, 63–71. (doi:10.1016/j.gene.2011.05.027)

86. Strick TH. 2019 Z. Phylogeny. In *Handbook of zoology, annelida, volume 1: annelida basal groups and pleistoannelida, sedentaria (eds G Purschke, M Böggemann, W Westheide)*, pp. 37–56. Berlin, Germany: Walter de Gruyter GmbH & Co. Kg.

87. Bleidom C. 2019 Recent progress in reconstituting lophotrochozoan (spiralian) phylogeny. *Org. Divers. Evol.* **19**, 557–566. (doi:10.1016/j.13127-019-00412-4)